VOLUME 105 PART B NUMBER 23 SEPTEMBER 1958

The Proceedings

OF

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871: INCORPORATED BY ROYAL CHARTER 1921

PART B

RADIO AND ELECTRONIC ENGINEERING
(INCLUDING COMMUNICATION ENGINEERING)

SAVOY PLACE · LONDON W.C.2

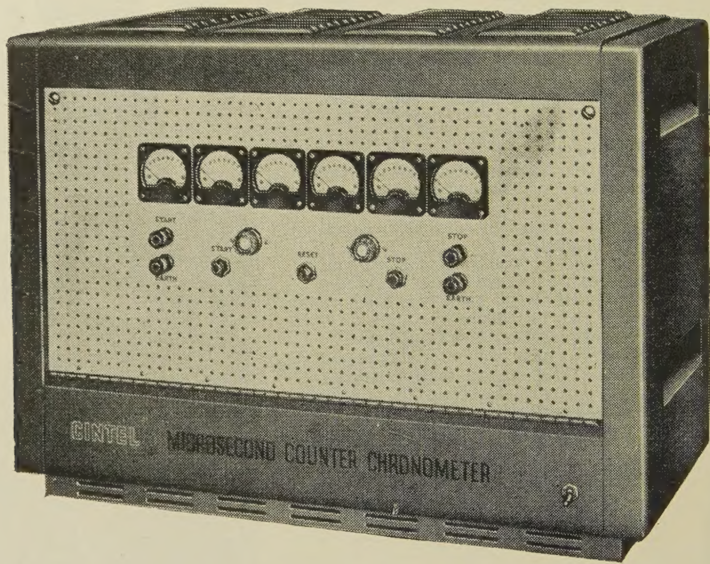
Price Ten Shillings and Sixpence

THE INSTITUTION OF ELECTRICAL ENGINEERS

FOUNDED 1871 INCORPORATED BY ROYAL CHARTER 1921

PATRON: HER MAJESTY THE QUEEN

MICROSECOND CHRONOMETER



This equipment is designed to measure small intervals of time to a high order of accuracy and two ranges are provided:

(1) 1 μsec. to 1 sec. in steps of 1 μsec.
(2) 10 μsec. to 10 sec. in steps of 10 μsec.

Accuracy of each range is better than ± 0.005% ± the step interval.

Full details on this and other 'Cintel' Chronometers are available on request.

CINEMA TELEVISION LTD

A COMPANY WITHIN THE RANK ORGANISATION LIMITED

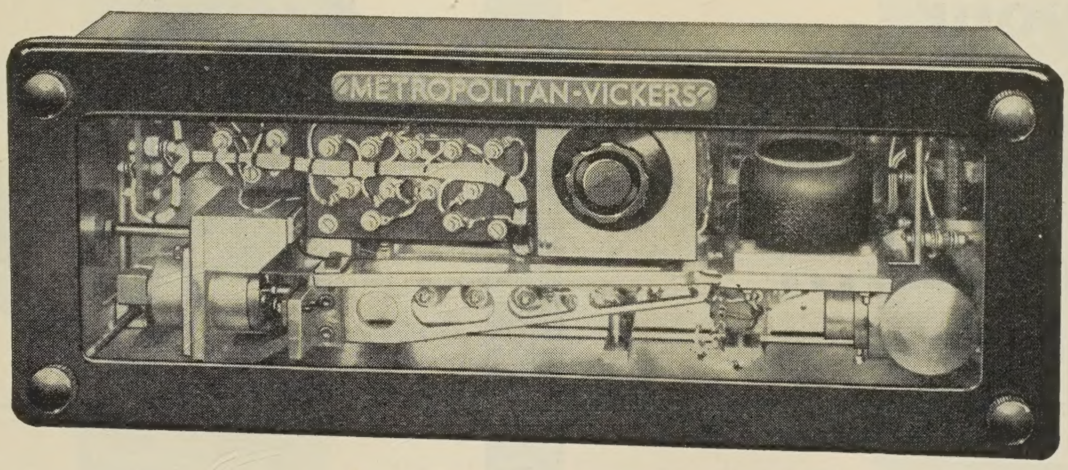
WORSLEY BRIDGE ROAD · LONDON · S.E.26
HITHER GREEN 4600

SALES AND SERVICING AGENTS:
Hawnt & Co. Ltd., 59 Moor St. Birmingham, 4
Atkins, Robertson & Whiteford Ltd., Industrial Estate, Thornliebank, Glasgow
McKellen Automation Ltd., 122 Seymour Grove, Old Trafford, Manchester, 16

METROVICK

Pneumatic-Electric Converter

Type P.C.8



the perfect link between pneumatic & electronic control systems

The Type P.C.8 Pneumatic-Electric Converter has been designed primarily as a link between pneumatic and electrical control systems where standard units in the two control mechanisms need to be combined. The input pressure is the normal 3-15 p.s.i. gauge used in process controllers and the output voltage (100 V max.) is suitable as, for instance, a reference voltage for most commercial motor speed controls. Other applications include conversion of air pressures for electrical recording.

Write for leaflet No. ES4566/2

METROPOLITAN-VICKERS
ELECTRICAL CO LTD TRAFFORD PARK MANCHESTER, 17

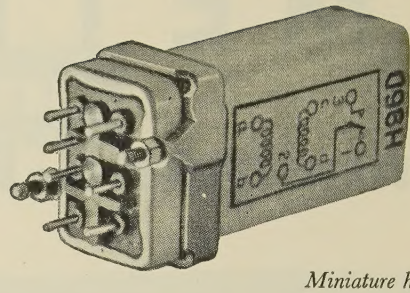
An A.E.I. Company

WINGS DAY
SEPT. 20th
GIVE FOR THOSE WHO GAVE

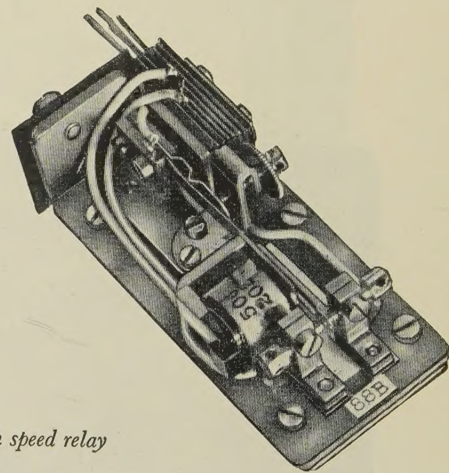
INDUSTRIAL PROCESS CONTROLS

R/E 704

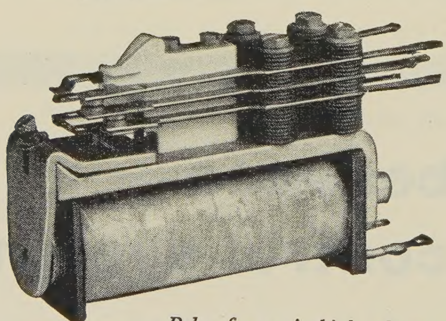
Prompt delivery of your RELAY needs...



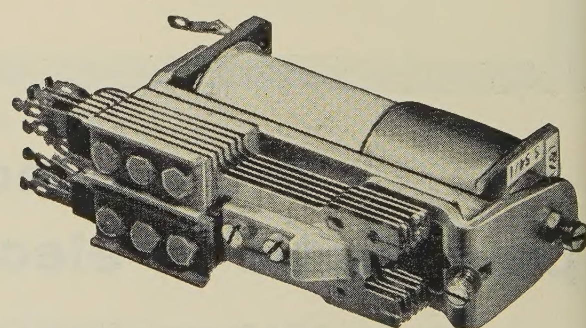
Miniature high speed sealed relay



High speed relay



Relay for use in high voltage and high frequency circuits



B.P.O. 3000 type relay

The versatility of the telephone type relay has led to its widespread use outside the telephone industry. Our Woolwich Works developed the original BPO 3000 type relay which has since won world renown. With unrivalled experience in the design of relays for a wide variety of special applications, our engineers can give individual attention to problems in this field. **Prototypes can be delivered almost immediately, with bulk supplies following in quick succession.** We would welcome your enquiries.

SIEMENS EDISWAN

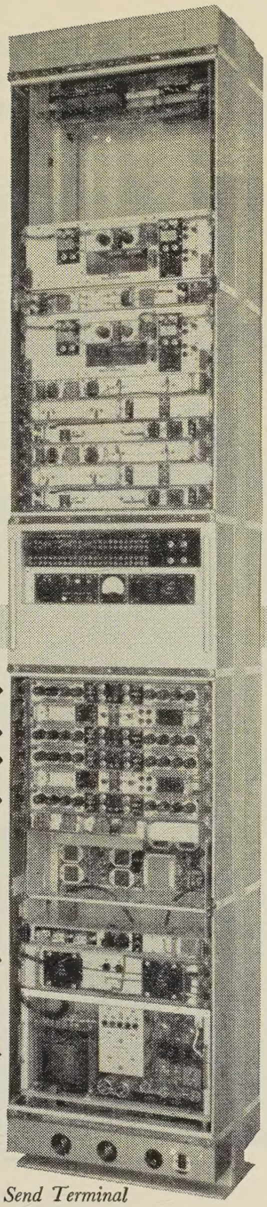
SIEMENS EDISON SWAN LTD
An A.E.I. Company
Woolwich, London S.E.18
Telephone: Woolwich 2020 Extn. 621

TAB 508

Programme Channel Equipment

CS12/CM

UNIDIRECTIONAL
OR BOTHWAY CIRCUITS
50 c/s—10 kc/s



Send Terminal

THESE ILLUSTRATIONS ARE OF A TYPICAL 4-CHANNEL UNIDIRECTIONAL SYSTEM

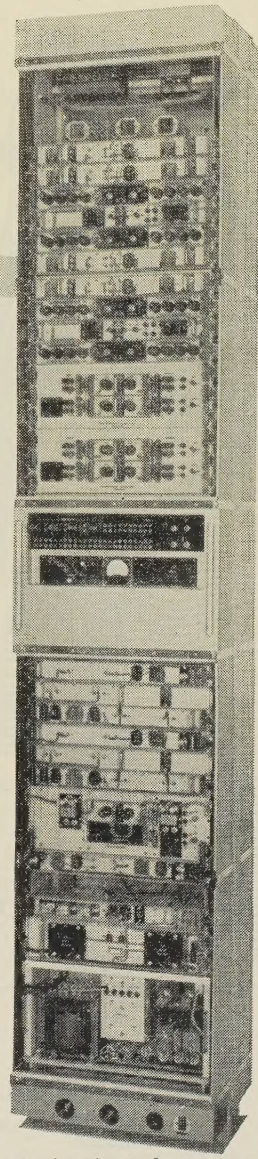
Wide range audio input levels
Output 24 kc/s—34 kc/s
or 84 kc/s—94 kc/s

The overall performance of three systems in tandem (6 terminals) is within the C.C.I.T.T. recommended limits for a 'normal' programme circuit.

Bays may be equipped for either unidirectional or bothway circuits, making the equipment suitable for both studio to transmitter circuits and national telephone network use.

Polyphase modulation is used and a small portable test set is supplied with the equipment for adjustment of the phase modulators.

A bothway channel is complete on one side of a standard 9ft. x 20½ in. rack with power supplies and control panel. The only wiring required to the main carrier system is for the carrier frequencies and input/output leads.



Receive Terminal

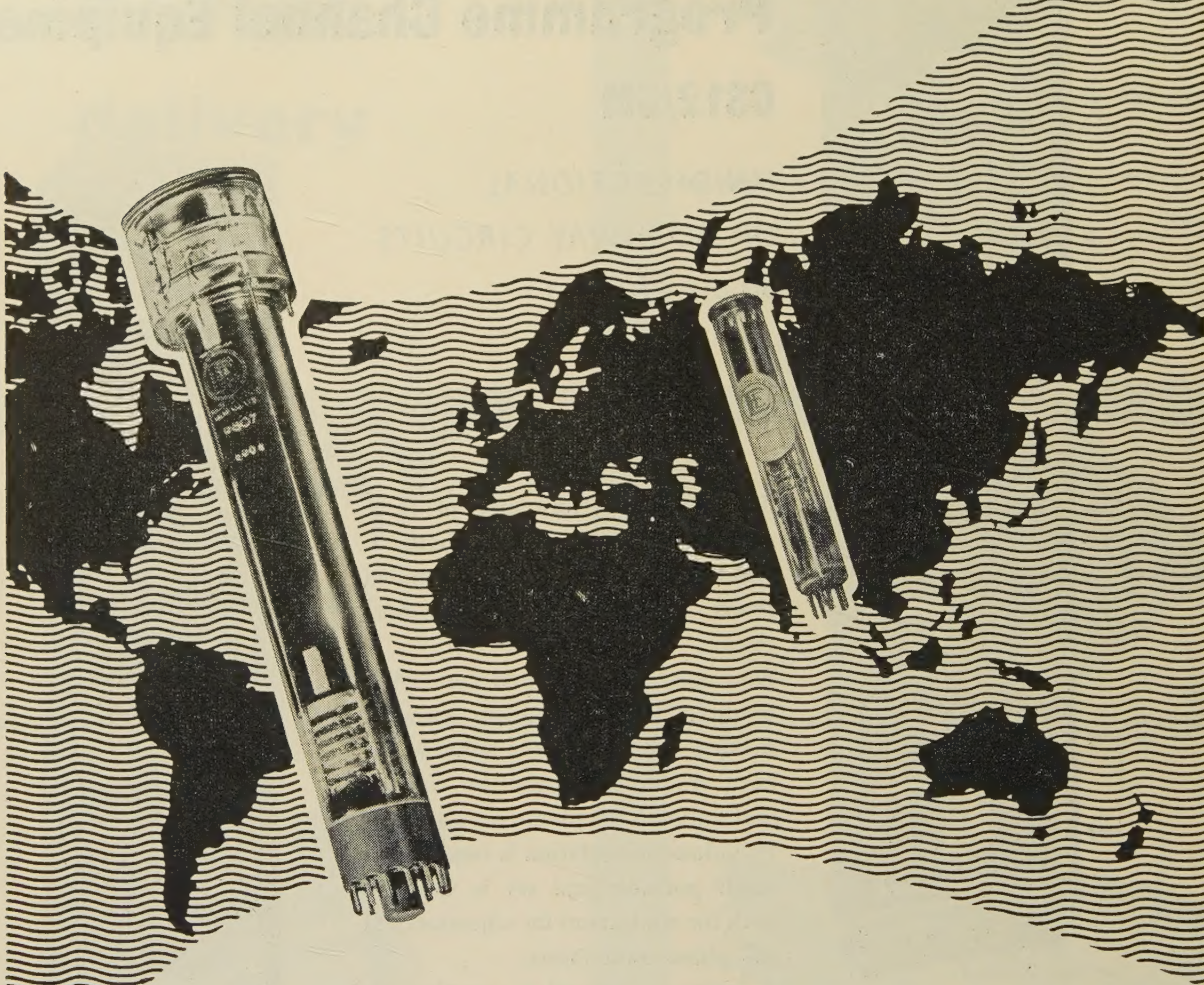
extending → **the frontiers of telecommunications**

SIEMENS EDISWAN

SIEMENS EDISON SWAN LTD *An A.E.I. Company*
Telecommunications Transmission Division, Woolwich, London SE18, England
Cables: Sieswan London

TAB 748

TELEVISION CAMERA TUBES



The following countries are some of those supplied by E.E.V. with Image Orthicon and Vidicon camera tubes:—

Australia	Iraq
Austria	Italy
Belgium	Norway
Canada	Poland
Cyprus	Portugal
Czechoslovakia	Russia
Denmark	Sweden
Finland	Switzerland
France	U.K.
Germany	Venezuela
Holland	Yugoslavia

The range includes:—

E.E.V. type	*American Equivalent*	*Description*
P.807	—	3″ Image Orthicon (field mesh type)
P.809	—	3″ Image Orthicon (field mesh type)
P.810	6198	Vidicon for industrial use
P.811	—	4½″ Image Orthicon
P.813	6326	Vidicon for film pick-up
P.816	5820	3″ Image Orthicon
P.817	6474	3″ Image Orthicon for colour pick-up

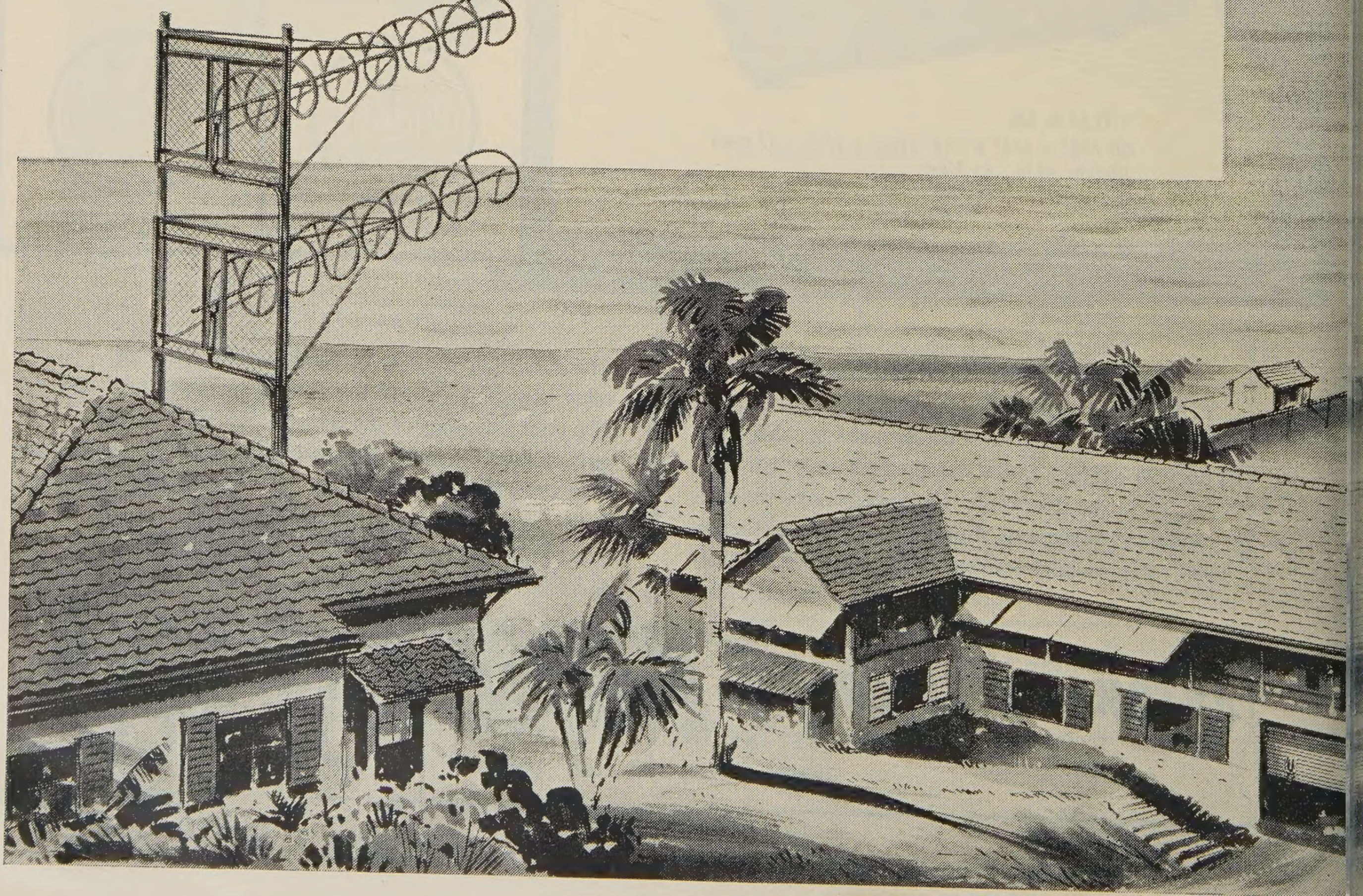
Write for full technical data of the complete range.

ENGLISH ELECTRIC

ENGLISH ELECTRIC VALVE CO. LTD.

Chelmsford, England
Telephone: Chelmsford 3491

AP300/108



p in touch...

will serve you – for less!

now available

G.E.C. 5-CIRCUIT JUNCTION RADIO EQUIPMENT

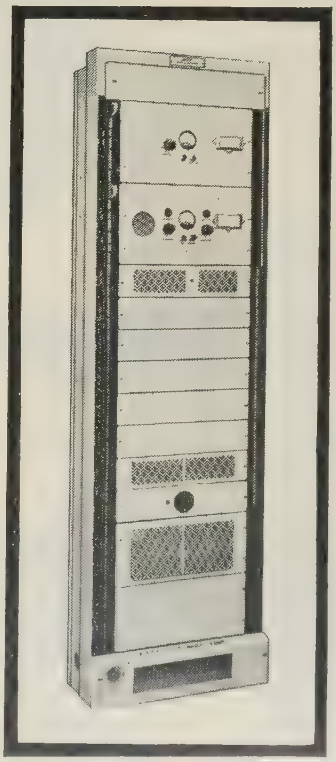
providing 5 speech circuits each having a bandwidth of

300 c/s to 3400 c/s

This operates over distances of up to 40 miles (65 km), without repeaters in any one of the following frequency bands:

71.5	**Mc/s**	**to**	**100**	**Mc/s**
132	**Mc/s**	**to**	**156**	**Mc/s**
156	**Mc/s**	**to**	**184**	**Mc/s**
235	**Mc/s**	**to**	**270**	**Mc/s**

Helical aerials having a forward gain of 13 db are available for the frequency band 156 Mc/s to 184 Mc/s in addition to the standard Yagi aerial.

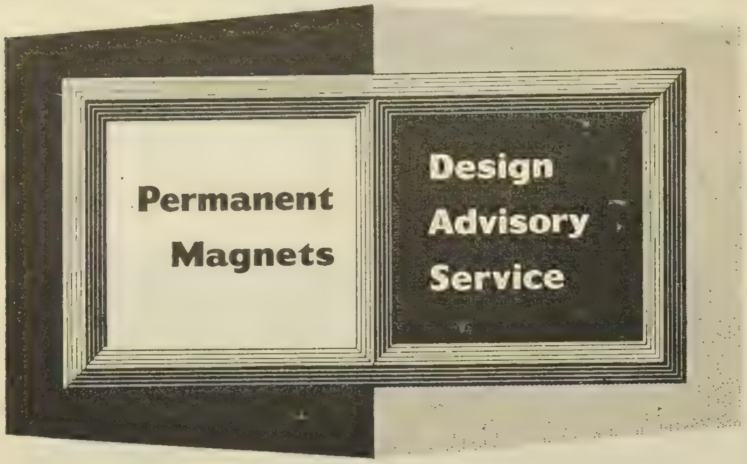


For further information, please write for Standard Specification SPO.5051.

THE GENERAL ELECTRIC COMPANY LTD OF ENGLAND
TELEPHONE, RADIO AND TELEVISION WORKS, COVENTRY

G.E.C.14


TYPE 2S005 N-P-N
SILICON GROWN DIFFUSED SWITCHING TRANSISTOR
TYPE 2S004
TYPICAL CHARACTERISTICS
SEMICONDUCTOR PRODUCTS
SEMICONDUCTOR APPLICATION REPORT
TEXAS INSTRUMENTS LIMITED
BEDFORD, ENGLAND



No.6

Methods of Magnetising Magnets

Advertisements in this series deal with general design considerations. If you require more specific information on the use of permanent magnets, please send your enquiry to the address below, mentioning the Design Advisory Service.

Permanent magnets require a magnetising force proportional to the coercivity of the material. It is extremely important that the magnetising force is not below the specified minimum value, otherwise a reduced performance may be obtained from the magnet.

To assist the designer these values are given below.

MATERIAL	MAGNETISING FORCE c.g.s.	MAGNETISING FORCE M.K.S.
'Ticonal' C, G, G_X and L	2,500 AT/cm.	0.25×10^6 AT/m.
'Ticonal' K	3,600 AT/cm.	0.36×10^6 AT/m.
'Reco' 3A	2,000 AT/cm.	0.20×10^6 AT/m.
'Magnadur' I	12,000 AT/cm.	1.2×10^6 AT/m.

Modern magnet materials require considerably more magnetising power than earlier materials. It may be as high as 40 times that required by tungsten or chrome steels. In order to obtain the maximum effect from the magnetising current it is recommended to short circuit the magnet during magnetisation by a heavy iron yoke.

The magnetising current may be obtained by several methods; some of which are outlined below.

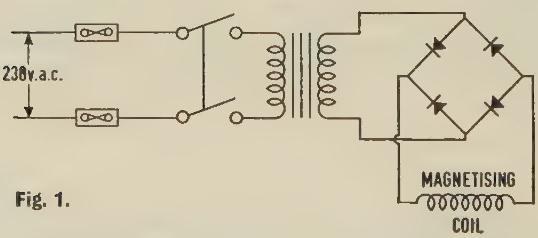


Fig. 1.

a. Metal Rectifiers (Fig. 1.)—generally preferable where it is possible to use a low current and many turns on a pre-wound coil. If an electromagnet is used it should be possible to attach specially shaped pole pieces to magnetise various types of magnets or magnet assemblies.

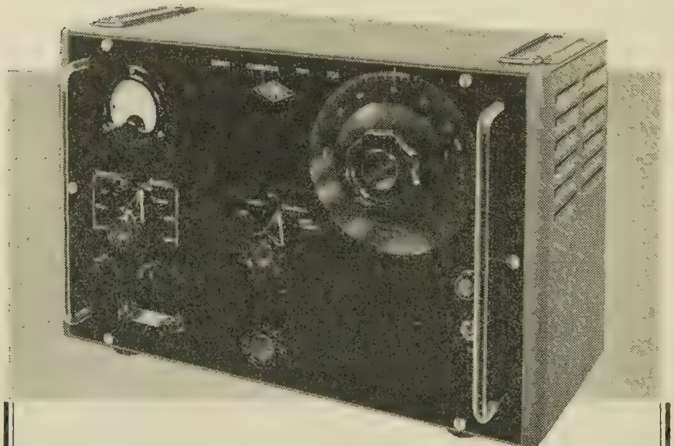
b. Ignitron Pulse Circuits—the use of ignitrons for controlling half cycle pulses is recommended where repetition magnetisation is required. The current available directly from the mains is usually sufficient for magnetising most magnets, but if a higher current is required, specially designed transformers are available for stepping the current up to higher values, of the order of 100,000 amps. This method is not recommended for magnets of large section as the eddy-current shielding effect produced during pulse magnetisation opposes the magnetising field and incomplete magnetisation may result.

c. Storage Accumulators—these are normally recommended for supplying current to coils for experimental work. When using accumulators adequate precautions should be taken for breaking the highly inductive circuit carrying high current.

d. Capacitor Discharge—the high current usually associated with magnetising can be obtained by discharging the current from a large charged condenser. This method generally uses an ignitron both as a switch and as a means to prevent oscillation with consequent partial demagnetisation of the magnet.

e. Motor-generators—can be particularly useful for supplying the current necessary for magnetising purposes. A wide variety of these are available to suit individual requirements.

SIGNAL GENERATOR
TYPE 702



The Airmec Low Frequency Signal Generator has been designed to provide a stable, accurate and easily controlled source of voltage in the frequency range 30 c/s–30 kc/s.

OSCILLATOR
A resistance-capacitance type of oscillator is employed to obtain a stable output waveform with very low harmonic distortion. Such disadvantages as spurious R.F. content in the output signal, and poor frequency stability, which are commonly associated with beat-frequency oscillators, are thereby eliminated.

OUTPUT CIRCUIT
The output circuit of the instrument uses a screened and balanced transformer to enable unbalanced, balanced and floating conditions to be obtained, and a 600 ohm constant impedance attenuator provides steps of 0, 20, 40 and 60 db of attenuation under all output conditions. A single level control is incorporated for setting the output to any desired value, and the amplitude is monitored by a diode valve voltmeter calibrated both in open circuit volts and in db relative to 1 milliwatt in 600 ohms.

BENCH USE and FORWARD MOUNTING
The instrument is enclosed in a standard Airmec case and is therefore both for bench use and for forward mounting on a 19-inch rack.

Immediate Delivery
Full details of this or any other Airmec instrument will be forwarded gladly on request.

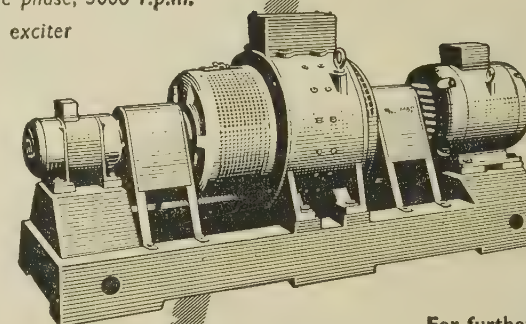
Airmec

INSTRUMENTS FOR RESEARCH & INDUSTRY

AIRMEC LIMITED · HIGH WYCOMBE · BUCKS
Telephone: High Wycombe 2060

ELECTRICAL EQUIPMENT

Motor Alternator Set
12 kVA, 400 c.p.s.
single phase, 3000 r.p.m.
with exciter



For further details write to the manufacturers.

NEWTON DERBY

NEWTON BROS. (DERBY) LTD

ALFRETON ROAD DERB

PHONE. DERBY 47676 (4 LINES) GRAMS· DYNAMO, DERBY
London Office: IMPERIAL BUILDINGS, 56 KINGSWAY W.C 2

ZENITH
(REGD. TRADE-MARK)

PHASE SHIFTING TRANSFORMER

This instrument provides co venient means for adjustin the phase angle or powe factor in alternating curren circuits when testing single o polyphase service meter wattmeters, or power facto indicators, etc. It is also th simplest means for teachin and demonstrating Alternatin Current Theory as affectin phase angle and power factor.

Illustrated brochure free on request

The ZENITH ELECTRIC CO. Ltd
ZENITH WORKS, VILLIERS ROAD, WILLESDEN GREE
LONDON, N.W.2

Telephone: WILlesden 6581-5 *Telegrams:* Voltaohm, Norphone, Londo

MANUFACTURERS OF ELECTRICAL EQUIPMENT INCLUDING RADIO AND TELEVISION COMPONENTS

Wire for electrical progress

Lewmex or Lewkanex, Lewcosol or Lewcoglass—just a few of the 'LEWCOS' range of insulated wires designed to meet every need or application in the electrical industry, backed by intensive research and a first class technical service.

LEWCOS

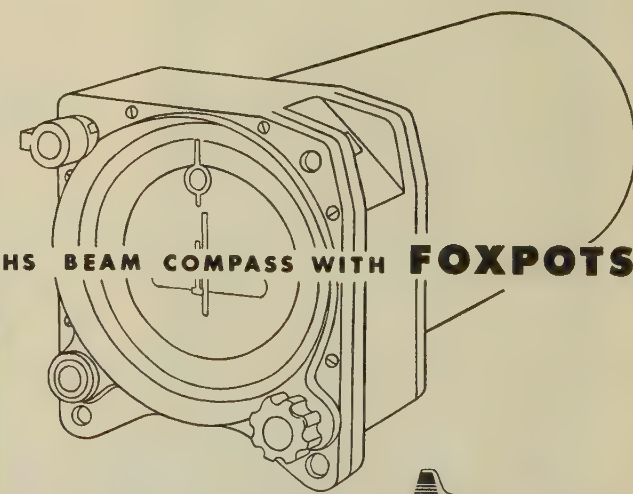
the largest manufacturers of insulated wires and strips in Europe

THE LONDON ELECTRIC WIRE COMPANY AND SMITHS LIMITED

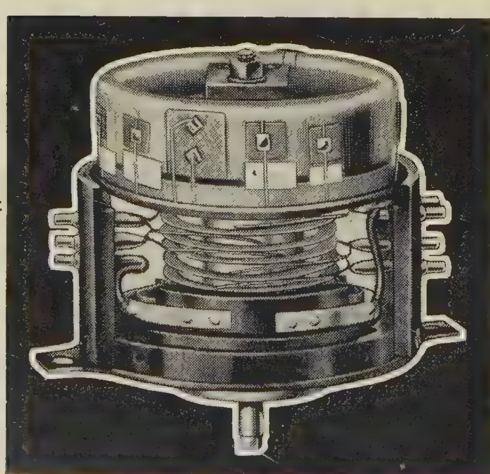
LEYTON · LONDON · E10

---

only the best for Britannia



SMITHS BEAM COMPASS WITH **FOXPOTS**



BRITANNIA AND OTHER AIRCRAFT RELY ON SMITHS

—SMITHS RELY ON FOX-POTS

FOX-POTS ARE ALWAYS ON THE BEAM

The Fox-Pot used in the Smiths Flight Beam Compass is a modification of a standard Fox Potentiometer.

Write for details of the full range of Fox-Pots to:—

P. X. FOX LTD., Dept. 215, HARKSWORTH ROAD, HORSFORTH, YORKSHIRE

P·X·F

"The Greeks had a word for it"!

ΒΑΡΛΕΫ
ΜΑΓΝΗΤΙΣ ΛΙΘΟΣ

WINGS DAY
SEPT. 20th
GIVE FOR THOSE WHO GAVE

Or, to be more precise, Varley Solenoids, which push, pull, punch or press from *fractions of an ounce to hundreds of pounds* through *thousandths of an inch to five inches.* Remote control by Varley Solenoids saves time and money. 100% British in design and manufacture. All normal voltages and ratings "off-the-shelf". For any specific application, prototypes—7-10 days, quantity production—3-4 weeks.

Varley REGD. TRADE MARK **SOLENOIDS**

FOR REMOTE CONTROL

For full details of Varley Solenoids mail this coupon:—

OLIVER PELL CONTROL LTD
Cambridge Row · Woolwich · London, S.E.18 · England
Cables: Varlymag, Woolwich · Telephone: Woolwich 1422

NAME ....

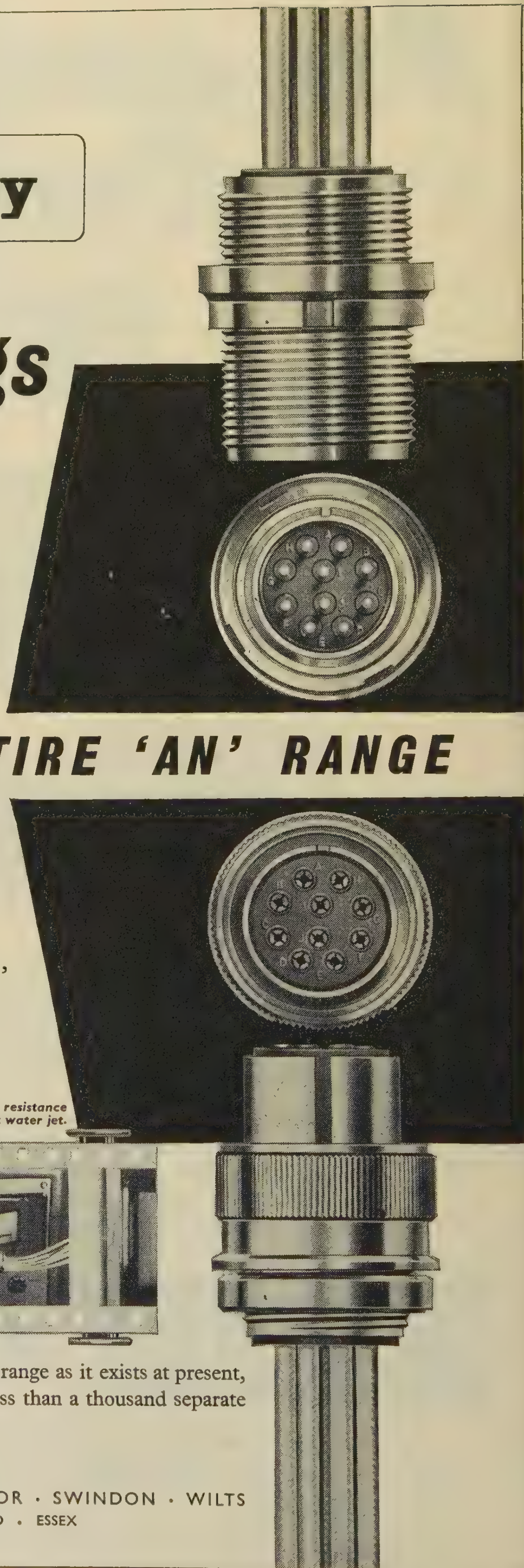
COMPANY ....

ADDRESS ....

IE.9. ....

OPS.3

A new Plessey range of Plugs and Sockets



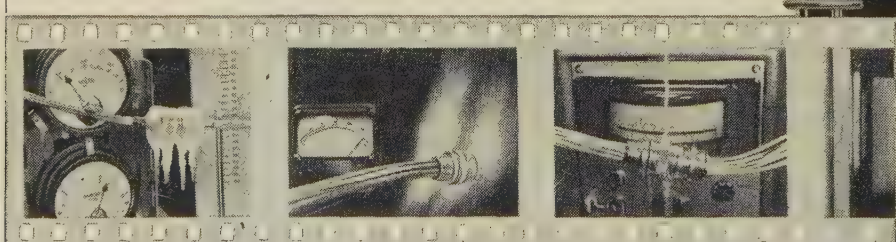
COVERING THE ENTIRE 'AN' RANGE

The Plessey UK–AN series of electrical connectors is now available and, for the first time from a non-dollar source, manufacturers will be able to obtain a full range of plugs and sockets completely interchangeable with the existing AN range.*

The Plessey UK–AN range has been designed and developed to M.O.S. Specification EL 1884 and RCS 321, and UK–AN connectors are fireproof, pressure sealed and environmental resisting. No separate wiring accessories are needed.

Write for test reports and full technical details.

Thawing out after low-temperature test at –60° C. *Fireproofness test (15 mins. at 1,100° C.)* *Testing insulation resistance under direct water jet.*



* There are many thousands of separate items in the AN range as it exists at present, whereas the ingenious Plessey UK–AN list comprises less than a thousand separate items, yet achieves the same service requirements.

ELECTRICAL CONNECTORS DIVISION

THE PLESSEY COMPANY LIMITED · CHENEY MANOR · SWINDON · WILTS

Overseas Sales Organisation: PLESSEY INTERNATIONAL LIMITED · ILFORD · ESSEX

PW2a

INDEX OF ADVERTISERS

THE PROCEEDINGS OF
THE INSTITUTION OF ELECTRICAL ENGINEERS

EDITED UNDER THE SUPERINTENDENCE OF W. K. BRASHER, C.B.E., M.A., M.I.E.E., SECRETARY

VOL. 105. PART B. No. 23. SEPTEMBER 1958

21.315.616 : 678.01

The Institution of Electrical Engineers
Paper No. 2616
Sept. 1958
©

THE FORTY-NINTH KELVIN LECTURE

'CHEMICAL STRUCTURE OF HIGH POLYMERS AND THEIR PHYSICAL BEHAVIOUR'

By H. W. MELVILLE, D.Sc., F.R.S.

(*Lecture delivered before* THE INSTITUTION, *6th February*, 1958.)

I should like to thank The Institution for the honour they ave done me in asking a chemist to give this, the Kelvin Lecture. must say I was somewhat hesitant at accepting the invitation ecause I could not claim to talk on any subject of electrical nterest or even on any topic that Kelvin illuminated by his enetrating and far-seeing mind. I would, however, like to try n the space of a lecture to look at the co-ordination of a number f sciences—in this case chemistry, physics and engineering—in he field of high polymers. New materials of any kind are of nterest not only to those who make them (in this case the hemist) but also to those who use them (the physicist and the ngineer). In fact there could not be a better modern example f the fusion of a number of separate sciences to produce a sum otal of knowledge that none could have achieved separately.

High-polymeric materials are not specially new, for rubber nd celluloid were discovered in the 19th century. The sys- ematic search for new materials of this kind for quite specific ses is, however, a much more recent development. The elationship between properties and chemical constitution has een one that has occupied attention for many years, and there s now beginning to emerge sufficient knowledge to approach hese problems in a much less empirical way than was possible ven twenty years ago.

The fundamental property of a high-polymeric material, as its ame implies, is its high molecular weight. This means mole- ular weights of upwards of 10 000 to values of the order of a illion for linear molecules and indefinitely large values for on-linear systems. Size is not the only thing that matters. he other essential feature is that the molecules should consist f long chains of atoms to which are attached additional atoms r small groups of atoms as side chains. We have therefore two in variables: the length and chemical nature of the chains mselves, and the chemical nature of the side groups. Both e a profound effect on the properties of the materials com- ising such molecules. Two general matters of principle are of importance here. For the sake of simplicity, assume that the chain wholly consists of carbon atoms with a single bond between each; in this case each atom with its associated groups can rotate with respect to the neighbouring atoms very freely. In the extreme case of a simple molecule like ethane [CH_3—CH_3] the speed of the CH_3 groups with respect to each other is about 10 rotations per second at room temperature. This corresponds almost to unrestricted rotation. If the temperature is raised or lowered the speed is increased or decreased not much more than proportionately with the absolute temperature. However, if bulky groups are attached to the carbon atoms the speed can be diminished and almost brought to a standstill. With this impediment to rotation the temperature coefficient of velocity of rotation is much increased, being exponentially dependent upon temperature. Similarly in the liquid and especially in the solid state, neighbouring chains and molecules also impede movement. This effect is likewise markedly affected by tem- perature. If, however, the carbon atoms are joined by double bonds, and more rarely by triple bonds, rotation is virtually impossible at room temperature. There are other structures, e.g. rings of atoms, which confer greater rigidity on an atomic chain. The result of this relative ease of rotation causes a normal type of chain to coil up into a random but continuously changing configuration. In a completely free chain, as would occur in an isolated molecule of polyethylene, the distance between the ends of the chain is given by $d = 2{\cdot}18n^{1/2}$, where d is the distance in ångströms and n is the number of carbon atoms in the chain, whereas the maximum outstretched length of the molecule is $1{\cdot}21n$. The ratio between the two is therefore $0{\cdot}53n^{1/2}$, which depends on the number of atoms in the chain. 10 000 such atoms in a chain is not a particularly large number, and therefore the ratio would be no less than 53 : 1. A coiled-up molecule will be generally spherical in shape: electron-microscope pictures of individual molecules confirm this supposition.

There are other faetors that control the shape of macro- molecules and again play a particularly vital part in determining the nature of the high polymer. The building units of polymers, i.e. the monomers, are normally liquids and sometimes solids.

r. (now Sir Harry) Melville is Secretary of the Department of Scientific and strial Research.

This means that there are considerable attractive forces between the monomers and hence also between the units of the polymers. Thus, if one imagines long chains of polymer molecules arranged side by side, the total force of attraction between the molecules can be very large indeed. These forces can be considerably increased by attaching to the chain of, say, carbon atoms, polar groups like $>C{=}O$, $=NH$, $—CN$, $—Cl$, as pendant structures. If the geometry of such groups is carefully arranged the attractive force can become very large. There is therefore a competition between these aligning intermolecular forces and the tendency for a molecule to coil up. Since the magnitude of the attractive forces is diminished by increasing temperature, the tendency will be for the molecule to coil up at high temperatures. This competition in fact determines whether a polymer will turn out to be an elastomer (a rubber), a fibre-forming material or a homogeneous supercooled liquid (a plastic).

The first stage in designing molecular structure to fulfil a particular role is to consider matters in this qualitative way. The more detailed modification of the structure to fit the exact role is the next step in the problem, for there are many mechanical properties of these three categories of substance which are affected markedly by the chemical nature of the backbone atoms and the pendant groups.

One might think a logical and qualitative approach along these lines would be the process of achieving the required properties in a relatively straightforward manner. (There is of course also the problem for the chemist of synthesizing the specified structures, but this is a matter which need not concern us here.) As in many scientific developments the result is in fact achieved by a combination of logical thought, the right hunches, acute observations at the right time and a very large amount of detailed, hard and patient experimenting.

It is convenient in this lecture, however, to consider the three classes of high polymers and to see how some of these generalized ideas can be used in particular examples.

RUBBERS

The real reason why rubber can be stretched to a far greater extent than any other solid is simply that in the undeformed state the molecular chains are coiled up. Stretching the rubber consists in the uncoiling of these chains under quite moderate tension. On release of the deforming force the rubber instantly returns to its former shape due to the thermal motion causing relative rotation of the carbon atoms to give the coiled-up configuration. In a good elastomer the ratio of the maximum outstretched length of the specimen to the normal length is as much as 8 : 1.

Normally one deals with vulcanized rubber, in which the rubber molecules are occasionally cross-linked together by suitable chemical processes. Unvulcanized rubber behaves quite differently. If stretched below 20° C it does not snap back to its original shape—it stays in a stretched configuration. In the stretched state, X-ray diffraction photographs demonstrate that the rubber has crystallized with the chains oriented parallel to the direction of stretch. It is easy to demonstrate mechanically that the rubber is anisotropic. If a small cut is made parallel to the direction of stretch it is easy to pull the rubber into two strips. If the cut is made perpendicular to the direction of stretch it is quite difficult to propagate the tear, or cut. These phenomena can only be observed below 20° C. If the stretched rubber is heated above 20° C it rapidly regains its original shape. Thus we can talk about the melting-point of rubber. Above it, crystallization and therefore alignment of chains just cannot occur. Next, if unstretched rubber is cooled to −60° C, it loses its rubber-like behaviour and becomes quite brittle like a supercooled liquid or glass. In the practical use of elastomeric substances some of these phenomena would be most undesirable. This crystallization has to be prevented so that the rubber can be used at temperatures much lower than 20° C. In practice this is mainly done by cross-linking, or vulcanization. Even so, at low temperatures natural and some of the varieties of synthetic rubber cannot be prevented from becoming brittle. This chemical cross-linking is not, of course, carried far enough to prevent the chains between the links coiling and uncoiling as the rubber is stretched and relaxed. Finally there is another important effect of cross-linking. If unvulcanized rubber were maintained in a stretched configuration the chains would tend to slip past each other so that eventually the rubber would acquire a permanently deformed state. This possibility of permanent deformation can be further prevented by reinforcing the rubber article with suitable fibrous materials, as is done, of course, in the ordinary motor-car tyre.

Rubber-like behaviour is a characteristic of hydrocarbon chains, but it is important that the geometry of the chains and associated groups is such that they do not too easily fit together so as to facilitate crystallization; otherwise there is, of course, no possibility of the chains coiling up. The importance of molecular geometry is strikingly illustrated in comparing gutta-percha and rubber. The configuration of their chains only differs in the disposition of the groups about the double bonds in the chain. Guttapercha is only elastomeric above 70° C.

These matters are illustrated in an even more striking way by the so-called polyurethane rubbers. Here one starts with a not so very large molecule, about 2 000 molecular weight (200 atoms in a chain, approximately), rather similar in chemical structure to polymers needed for making man-made fibres, where the prime necessity is that the molecules will fit together into ordered arrays of atoms as in a crystal. At this stage, however, the chains are not really long enough (200 Å) to form crystallites, which in fibres are usually of the order of 500Å. These molecules may, however, be linked together to give linear molecules by comparatively large and bulky groups of atoms, which make it quite impossible, because of the random distribution in the chains, for this crystallization to take place. Each individual chain is flexible and has a strong tendency to coil up so that an elastomeric material is the result. By a further modification of the synthetic process cross-linking can also be achieved.

All the commonly used synthetic rubbers and, of course, natural rubber, exhibit elastomeric behaviour over a limited temperature range. There is naturally a need to extend this range, and extending it downwards is possible by a number of methods. One consists in adding small molecules as plasticizers or molecular lubricants, to prevent the rubber forming a glass at these low temperatures and facilitating the movement of the rubber molecules with respect to each other.

A recently discovered method of preventing crystallization* consists in making a slight chemical modification to the rubber. In the presence of a compound—a thiolacid containing the group COSH—the disposition of the groups around the double bond connecting two of the central atoms in the unit of the rubber molecule can be altered. The result is that the regular molecular geometry is disturbed, and it then becomes impossible for the chains to form crystallites. The striking effect of this small change in molecular structure is shown most clearly (Fig. 1) by stretching the rubber to double its original length and measuring the stretch after a period of one hour. At −30° C ordinary rubber has recovered to only 7% of its original length; on the other hand, the modified rubber has recovered to 98%.

Extending the temperature range to higher limits means giving the molecules structures altogether different from the

* British Rubber Producers Research Association.

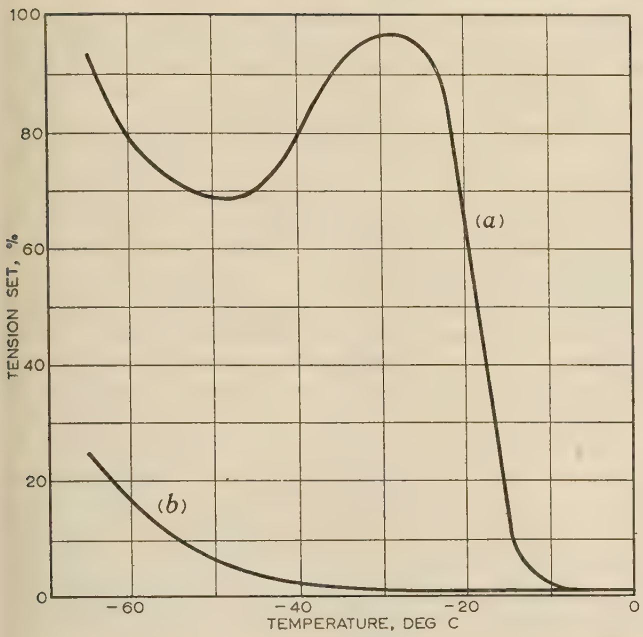


Fig. 1.—Effect of temperature on behaviour of normal and anticrystallizing rubber.

(*a*) Recovery of anticrystallizing rubber.
(*b*) Recovery of ordinary rubber.

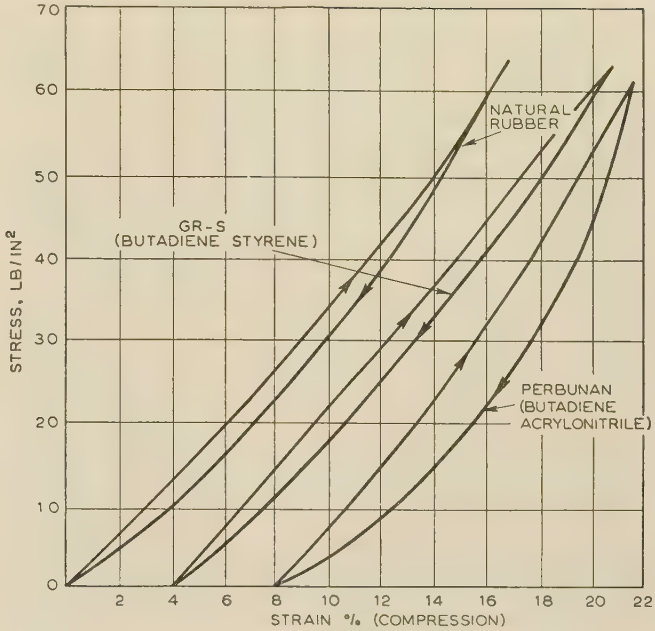


Fig. 2.—Stress/strain hysteresis loops for a variety of rubbers.

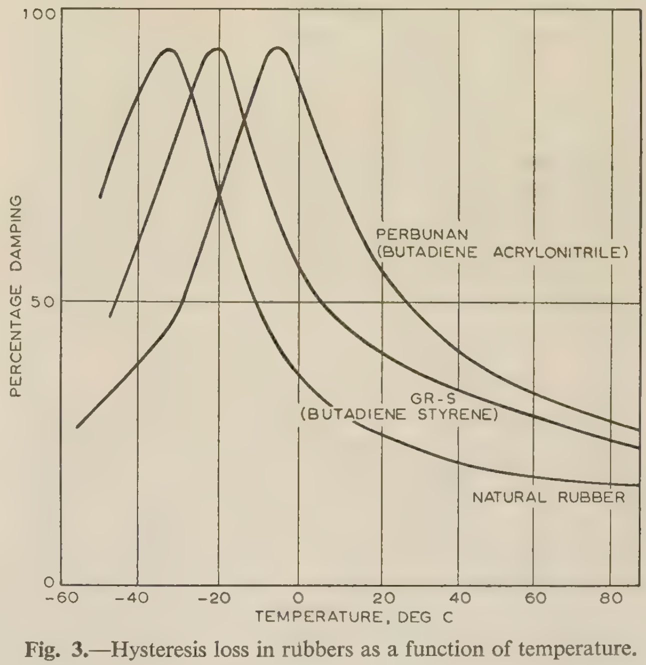


Fig. 3.—Hysteresis loss in rubbers as a function of temperature.

carbon backbone, and in fact it is necessary to consider a backbone of silicon and oxygen atoms in repeating units with various groups attached to the silicon atom (e.g. CH_3). Physically, these polymers—the silicones—are substances ranging from mobile liquids to rubbers, depending only on the molecular weight of the molecules comprising them. The liquid members of this series are far more compressible than ordinary liquids, but more extraordinary still is the fact that the viscosity (the intermolecular forces) is not affected by temperature to anything like the extent of liquids such as hydrocarbons. The overall temperature dependence extends the usefulness of the silicon rubbers over a much wider range of values than the ordinary hydrocarbon rubbers.

By its very nature, rubber is liable to be flexed and nowhere more often than in motor-car tyres. Stress/strain curves (Fig. 2) for rubber are quite unlike those for other solids (apart from the fact that the strain is very much larger before rupture occurs). There is considerable hysteresis in going round a stress/strain cycle. On relieving the stress, the energy not recovered, as measured by the area of the loop, appears as heat. Repeated cycles in such a bad conductor of heat as rubber can soon raise its temperature to such high values that other properties of the rubber, e.g. abrasion resistance, are seriously diminished. Hysteresis loss varies a good deal among rubbers: natural rubber exhibits the least hysteresis, some of the common synthetics come next and butyl rubber exhibits the greatest. Further, these effects are temperature sensitive, and at 100° C, for example, the differences largely disappear (Fig. 3). This is an effect closely influenced by molecular geometry, but no rules have yet been evolved that make it possible to determine what structure will give rise to elastomers having a minimum hysteresis.

In the transmission of power to road wheels this is also important and high hysteresis can affect adversely the performance of an automobile.

Apart from these matters, abrasion resistance is one of the most sought after characteristics of rubber—for its largest use is in tyres. This much, however, is known about abrasion. It does not consist in the chemical breakdown of rubber molecules under mechanical strain into volatile compounds. It appears that small pieces of rubber are removed from the specimen mechanically. What probably happens is that an initial cut is made; under stress the cut is propagated, and it may be that the propagation is the rate-determining factor in the whole process of abrasion; finally a piece of rubber becomes detached from the specimen. Chemical structure has a most important effect on crack propagation; for example, the polyester rubbers are outstanding in their resistance to it. However, natural and to a greater extend G.R.S. rubbers would show up very poorly in this respect were it not for the introduction of the filler, carbon

black. And even the nature of the carbon black has a profound effect on abrasion resistance of the composite material. Here we have a heterogeneous system which is much more difficult to deal with scientifically. It would seem that one could visualize the problem in this elementary kind of way. The rubber at the apex of the incision is under considerable tension, and the tendency would be for the molecules to line up perpendicular to the direction of propagation of the crack, thereby inhibiting the propagation. With a non-crystallizing rubber like G.R.S., crack propagation is easy. Part of the function of carbon black is to prevent crack propagation by interposing an obstacle.

As has been mentioned, the elastomeric state of matter is the result of the balance of forces between coiling-up and intermolecular alignment. If one were approaching molecular design logically, one would therefore keep the latter interaction at a reasonable minimum so as to allow the other force to play a dominating role. The question now arises as to what kinds of solid result when the aligning force becomes predominant. In fact these are the plastic and fibrous materials. There is a large group of plastic materials—polystyrene, the polymethacrylates, polyvinylacetate, polyvinylchloride—where the chemical structure is such—all the units contain polar groups—that the intermolecular forces are quite high. But their molecular geometry is not especially symmetrical and the polar groups are fairly bulky. The point about symmetry is that in these structures there are groups attached to alternate carbon atoms in the chain. The disposition of these groups on either side of the chain is purely random. This untidy molecular pattern makes it impossible for the chains to align themselves precisely so that intermolecular forces can come into full play. These intermolecular contacts will be at random too. The result is that the solid does not exhibit elastomeric properties and is physically entirely isotropic and in many cases is a glass-clear material. Because of the relative infrequency of interchain contacts the softening point is rather low, but this means that the substance will flow under pressure at moderate temperatures and so is ideally suited to the manipulative procedures used in the plastics industry. In this class of substances it is extremely difficult to make any attempt to correlate mechanical properties with chemical constitution.

CRYSTALLIZING HIGH POLYMERS

There is a third class of high polymers where it has been possible to control the internal molecular geometry fairly exactly, and more recently methods have been devised for getting an extremely precise control comparable, in fact, to that which seems to be achieved in the synthesis of naturally occurring large molecules like proteins. Thus, not only are the polar interacting groups suitably chosen, but their spacing along the chains can be controlled and the disposition of the groups with respect to the chains also determined. In this way interchain attractions can be brought into full play. In general, this means that the tendency to alignment is strongly displayed, and these substances naturally show a strong tendency to crystallize and to exhibit relatively high melting-points.

The simplest of these substances is polyethylene. When prepared at about room temperature this is simply a linear chain of carbon atoms with hydrogen atoms attached. On the other hand, when prepared at high temperatures (200° C) and high pressures (1 000 atm) the molecules have many branches attached to the linear backbone. This difference in structure has important effects on properties. The similarity of the material to paraffins is perhaps shown most clearly when the specific heat of the high-pressure variety of polyethylene is plotted as a function of temperature. Over a very limited temperature range the specific heat rises to indefinitely large values, showing that the material must melt within a small range of temperatures. The liquid so obtained thus has a specific heat similar to that of liquid paraffins. When the liquid is quenched rapidly to room temperature the density of the solid is 0·91. On the other hand, when cooled very slowly to the same temperature, the density is much higher, 0·98. X-ray diffraction data show that the latter material is polycrystalline, the former amorphous. At room temperature the rate of crystallization is very small. The important point of principle here is that these polymers can thus be subjected to a measure of heat treatment—in a way similar to that of a metal—so that molecular arrangements can be altered to suit the purpose for which the polymer is required. A qualitative indication of the effect of this factor, as well as that of molecular weight, is shown in Fig. 4. Maybe the most

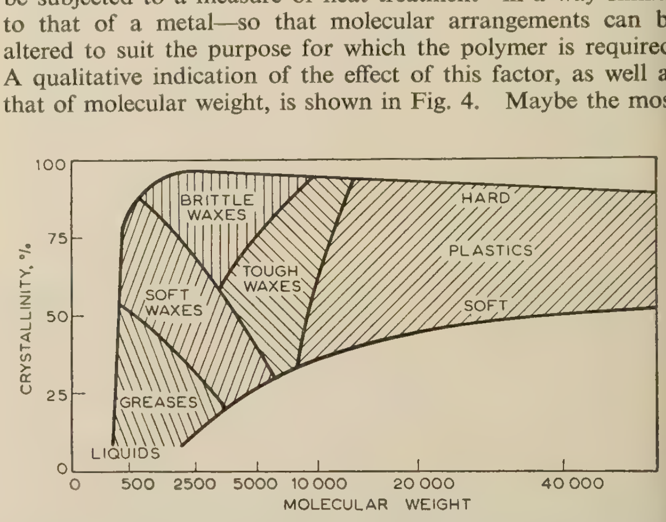


Fig. 4.—Properties of polyethylene as a function of molecular weight and crystallinity. [Richards, *Journal of Applied Chemistry*, 1951, **1**, p. 371.]

interesting region is around molecular weights of about 10 000, where the percentage crystallinity has a most profound effect on the nature of the product. The crystalline state is naturally the more thermodynamically stable, and therefore in considering the application of these materials in practice it is important that the transition does not occur during their working life. Many of the physical properties, e.g. softening-point under load, Young's modulus in tension, bending modulus and surface hardness, are likewise dependent on the crystalline/amorphous ratio.

The other variable at the disposal of the molecular designer is the amount of branching. In Fig. 5 is shown the effect of branching on the percentage of amorphous material in the solid. Similarly the melting-point of the linear material is about 20° C higher than that of the branched samples.

The final modification that can now be readily made to any variety of polyethylene to increase its melting-point almost indefinitely consists in exposing it to high-energy radiation, such as gamma radiation. One of the chemical changes is that hydrogen atoms are removed from the chains leaving free radicals behind. These radicals join up to give cross-links, and these give rise to a three-dimensional structure which at high enough temperatures decomposes rather than melts.

POLYPROPYLENE

A slight chemical modification to the polyethylene structure consists in replacing one hydrogen atom on each alternate carbon atom with a CH_3 group. Chemically this is done by polymerizing propylene $CH_3—CH=CH_2$. Here a more subtle chemical matter comes in which affects the properties of the polymer enormously. There are two kinds of geometrical structure possible: (*a*) one in which the methyl groups are all on

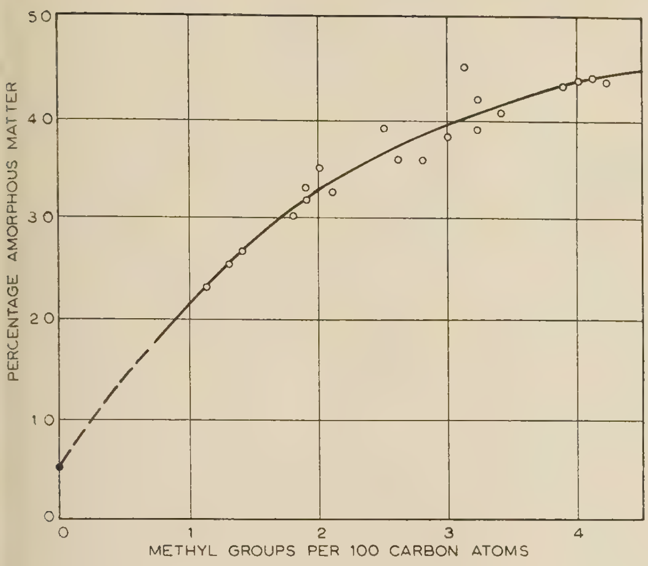


Fig. 5.—Effect of branching on amorphous content. [Raff and Allison, 'Polyethylene' (Interscience Publishers, Inc., New York).]

○ Polyethylenes.
● Polymethylene.

one side of the backbone of carbon atoms, and (*b*) the other in which the arrangement is purely random on both sides of the backbone. The (*b*) type gives rise to amorphous polymers of low melting-point, whereas the (*a*) type gives rise to crystalline material with a melting-point that, for a hydrocarbon, is very high (about 160° C). This is due entirely to the regular structure, which permits the chains to fit together into highly ordered regions characteristic of crystalline matter.

FIBRES

The mechanical properties of a man-made fibre can be set down fairly easily. It must have a high tensile strength (natural fibres have strengths up to 10 grammes per denier, or 100 kg/mm^2, a sufficiently high modulus of elasticity and an ability to extend to some extent without fracture; it must have a high melting-point—certainly above 200° C—for a variety of reasons; it must be insoluble in water, yet capable of absorbing water. Likewise it must have a high abrasion resistance, and be colourless and resistant to degradation by light and oxidation. In addition, it must have a variety of rather ill-defined or even difficult-to-define properties which contribute to the quality of the textile into which it is woven. Draping qualities, the so-called 'handle', and a variety of other properties come into this discussion.

Such a long list of requirements soon singles out relatively few structures that meet these needs. There are, however, some general properties which define the line of approach. In order to obtain high tensile strength along the length of the fibre, it is necessary that the molecules should be oriented in that direction and remain so during the lifetime of the fibre. This means that the polymer must be capable of crystallizing, and furthermore the molecules must be capable of being oriented parallel to the length of the fibre. In turn this restricts fibre-forming molecules to highly regular structures without having any bulky side chains attached to them. In attempting to achieve these objectives it is important that the substance should not be brittle like glass, otherwise continual flexing would soon reduce the tensile strength of the yarn to a very small value.

It is a curious fact that among the strongest natural and synthetic fibres the tensile strength comes out to be about 10 grammes per denier. In spite of the most strenuous efforts it is difficult to exceed this value. What is even more surprising is that this limited value seems to be independent of chemical constitution. From this it might be concluded that the figure must represent the ultimate strength of this kind of matter, i.e. the strength of the chemical bonds comprising the fibre, the assumption being made that, when the fibre breaks, the chemical bonds themselves are broken. But, in fact, fibres have only about one-hundredth of the strength calculated in this way. The other possible assumption about the way in which the fibre breaks under tension is that the chains slide past each other and the molecules or polymer remain intact. Even then actual strengths are less than theory predicts. These facts have been known for 20 years, but no rational explanation has yet been found. It is therefore tempting to speculate that much stronger fibres may yet be produced.

On the other hand, the modulus of elasticity approaches closer to theoretical expectations, and so it is unlikely that any large advances will be made in this property if fibre development continues.

The melting-point of a fibre is of vital importance. It must not be too low because, in general, fibres have to be washed in relatively hot water and they must be ironed to smooth out creases in the fabric. A reasonable minimum is 200° C, and 250° C is much more preferable. Of course, some fibres like cellulosic ones do not melt at all. Too high a melting-point is equally troublesome because many of the man-made fibres are not soluble and have to be spun by melting the material and extruding it through fine orifices. The melting-point is actually determined by two factors. At this temperature, T_m, the thermodynamic potentials of the solid and liquid phases are equal, and hence the second law of thermodynamics leads at once to the relationship $T_m = \Delta H/\Delta S$, where ΔH is the heat of fusion of the solid and ΔS the entropy change in the fusion. The former factor depends on the intermolecular attractive forces in the crystal and in the molten state. The latter introduces the idea of molecular shapes in the two phases. The larger ΔS, the lower is the melting-point. In the crystalline state the entropies of long-chain molecules probably do not differ greatly, but in the molten state the entropy will depend on the number of different configurations which the molecule can assume. If the molecule is rigid owing to the existence of certain types of chemical structure in the units, then the number of configurations in the liquid phase will be reduced. Hence the entropy of fusion will be small and the melting-point correspondingly high. For example, with Terylene the heat of fusion is 2·2 kcal per repeating unit, and for polyethylene adipate, 3·8 kcal. Terylene has a more rigid acid part to the molecule than the adipate. One might therefore expect the melting-point to be much higher for the adipate, whereas in fact it is only 58° C for the adipate and as much as 265° C for Terylene. The explanation is in the low entropy of fusion for Terylene at 4·0 cal per deg C, compared with no less than 12 cal per deg C for the adipate.

There are other geometrical factors that control melting-point and also tensile strength. So far as intermolecular attractions are concerned, in the molten state the molecules are to some extent free to assume configurations that permit these forces to be exerted to the full. In the crystalline state geometrical factors are much more important. In the nylon type of molecules the extra attraction between the chains is mainly due to what is known as a hydrogen bond. An essential element in the nylon structure in the group

```
  O   H
  ‖   |
—C—N—
```

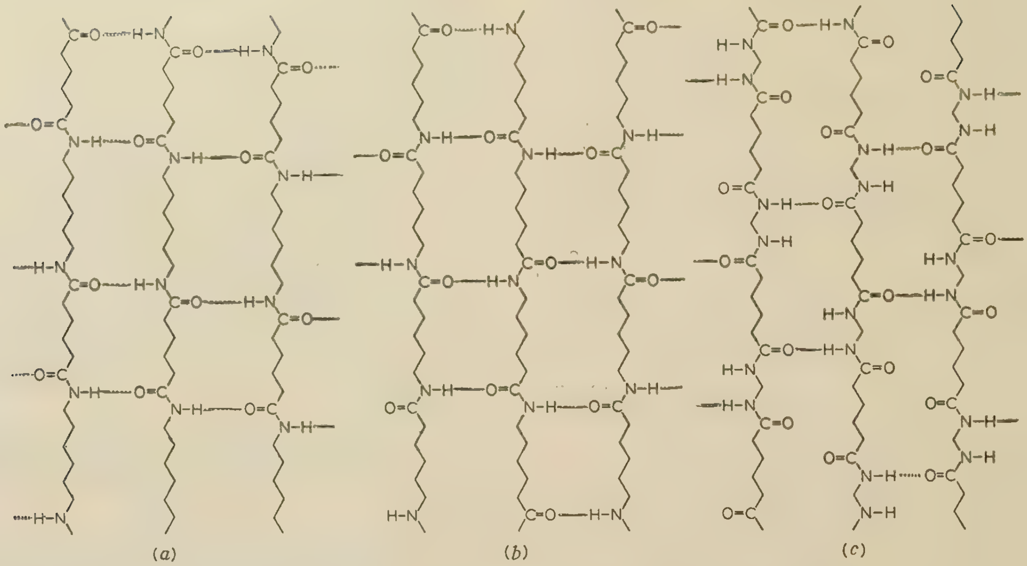


Fig. 6.—Hydrogen bonding in a variety of polyamides—nylon-type fibres. [Huggins, M. L., *Journal of Chemical Education* (Ohio), 1957, **34**, p. 480.]

(*a*) Hydrogen bonding in nylon 66 [Bunn and Garner].
(*b*) Hydrogen bonding in nylon 6 (polycaprolactam) [Holmes, Bunn and Smith].
(*c*) Proposed pattern of intermolecular hydrogen bonding in nylon 16.

This is spaced regularly along the chain. The hydrogen atom of the NH group has a strong tendency to interact with the oxygen of the CO group in a neighbouring chain (*a*). By slightly displacing one chain with respect to the next one, all these interactions are possible, as shown in Fig. 6. This accounts to a large extent for the high tensile strength of the nylon fibre. Again, in another version of the structure where the groups are separated by five carbon atoms, complete interaction is still possible (*b*). But in a chemically similar material (*c*) where the spacing is somewhat different, with only one carbon atom between the nitrogens, this pattern is not possible and some of the NH–CO groups simply cannot interact with each other, at any rate in one plane, and the melting-point of the material is correspondingly smaller than that of the other two. This is better shown (Fig. 7) when the melting-point of a series of such materials is plotted as a function of the number of carbon atoms between the NH–CO groups. Naturally there must be a gradual decrease, because in the limit the material would become polyethylene, melting at around 120° C. In fact, the melting-points are highest for the even number of carbon atoms because in these structures the geometry of the chain assures the maximum possible number of intermolecular interactions by means of this hydrogen-bonding effect.

Following this hypothesis to a further stage, there is another way of reducing interaction energy. This consists in chemically preventing hydrogen bonding between chains, and it is most easily done by replacing the hydrogen in an NH group by a CH_3 group. This does not contribute to interchain forces, since the hydrogen atoms do not interact with the CO group in the same way. This has the immediate effect of reducing the tensile strength of the material to almost impossibly small values, at any rate for a fibre. But these modified fibres exhibit a distinct property of very great elongation and behave, in fact, almost like elastomers in this respect. Thus the relationship between a fibre and a rubber is really here a very close one indeed and is influenced by relatively small changes in the chemical structure of the molecules comprising these materials.

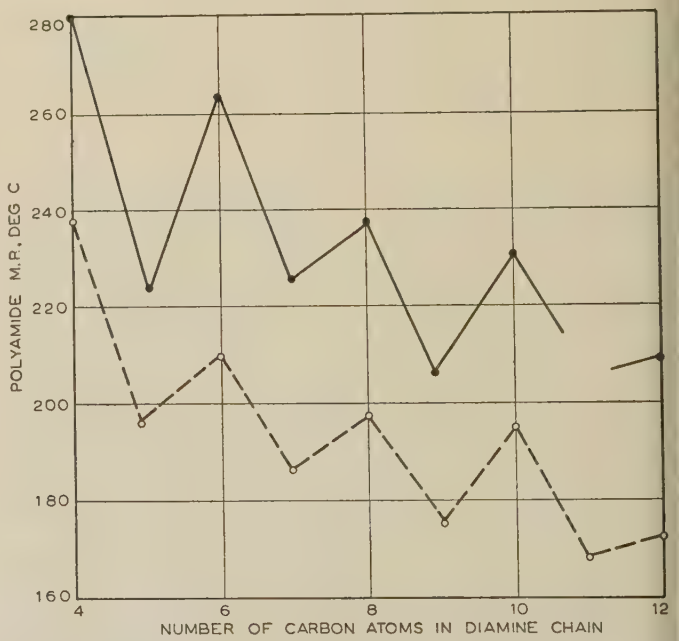


Fig. 7.—Melting points of polyamides from aliphatic diamines. [Hill, R.: 'Fibres from Synthetic Polymers' (Elsevier Publishing Co., Amsterdam).]

—— Polyamides from adipic acid.
---- Polyamides from sebacic acid.

ION-EXCHANGE FIBRES

One of the most interesting kinds of synthetic fibres is one resembling, in mechano-chemical behaviour, animal muscle; i.e. as a result of chemical reaction, mechanical work may be directly obtained from such a system. The precise chemical structure of the fibre is not known, but it is made by mixing high-polymer

molecules containing acid groups (COOH) with those containing alcohol groups (OH). The mixture, each component of which is soluble in water, can be drawn into fibres. These, when heat-treated, become insoluble in water owing to the formation of cross-links of the ester type formed by the introduction of the COOH and OH groups. The fibres are in effect solid acids. If such a fibre is suspended by a weight in an acid medium, the fibre swells and comes into equilibrium with its surroundings at a defined length. The fibre acid is weak, and therefore little dissociation of the acid occurs at the surface. If the acid medium is now replaced by an alkaline medium, i.e. one containing hydroxyl ions, these will react with the H^+ ions attached to the outside of the fibre to yield water and the fibre will be left with a negative electric charge. These charges cause the molecules to stretch by electrostatic repulsion. Finally, the fibre can be brought back to its original state by placing it in an acid bath. Thus mechanical work is done as a result of the neutralization of an acid by a base with the production of water.

The same result may be brought about by replacing one ion with another on the surface of the fibre, e.g. Na^+ ions and Ba^{++} ions. The extension of the fibre in solutions of equal osmotic pressure depends actually on the nature of the ion. This then may be made the basis of a thermodynamical cycle, as shown in Fig. 8. AB and CD represent two force/elongation curves for solutions of different composition. Along AB the fibre is in equilibrium with the surrounding solution and similarly along DC with another solution with an Na : Ba ratio of 10 : 1. Suppose the cycle is started at A, at which point the fibre is unstretched. Here the fibre contains mainly sodium ions on its surface. The fibre is then stretched isothermally and reversibly along AB. Work is done and as a result some of the Ba^{++} ions are transferred to the surrounding solution. Now, at constant force the fibre is transferred irreversibly to the bath with the higher Ba^{++} content. The fibre takes up the Ba^{++} ions, contracts and thus performs work. In the third step, CD, the stretching force is reversibly diminished to zero, when further work is done and more Ba^{++} ions are removed from the surrounding solution. The fibre is finally transferred to the original bath. Here the Ba^{++} ions are replaced by Na^+ ions, owing to the higher Na^+ content of the solution, and the fibre expands once again to point A. The net gain of work is shown by the shaded area in the diagram. It is, of course, done at the expense of transporting Ba^{++} ions from one solution to the other and Na^{++} ions from one solution to the other in the opposite direction. The actual efficiency of this mechano-chemical system is very low, about 1%. However, no real attempt has yet been made to get the optimum from such fibres for this kind of engine, and no doubt a much better performance could be achieved by such an effort. After all, the first steam engines were not very efficient, and even the modern railway locomotive operates at an efficiency of only about 5%.

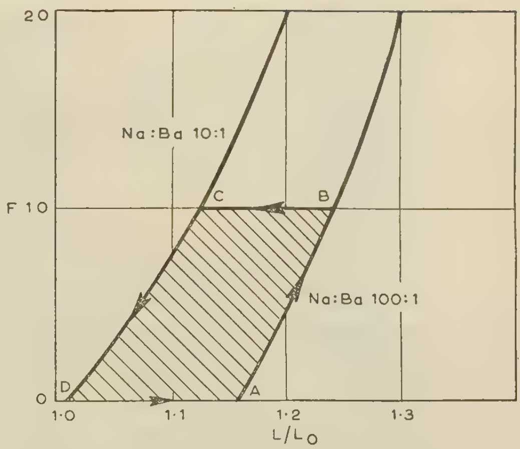


Fig. 8.—Mechano-chemical work cycle for a sodium-barium exchange on fibres of krilium-polyvinyl alcohol. [Katchalsky and Zwick, *Journal of Polymer Science*, 1955, **16**, p. 221.]

621.317.2

The Institution of Electrical Engineers
Paper No. 2682 M
Sept. 1958

INDUSTRIAL ELECTRICAL MEASURING INSTRUMENTS

A Review of Progress

By F. R. AXWORTHY, Member.

(1) INTRODUCTION

The past ten years have been a period of continuous development in the design, manufacture and application of industrial electrical measuring instruments. New and better materials are continually appearing, and these, coupled with the development of improved manufacturing techniques, have resulted in quite radical changes in the design of some types of instrument. Experimental studies of methods of displaying variable information have contributed towards providing instruments which are easy to read accurately and not unpleasant to look at. Since the war there has undoubtedly been a greater awareness amongst manufacturers that accurate measurement is one of the most important factors pertaining to a consistent quality of product. Scientific instruments are displacing approximate methods in more and more industries. This, of course, is a logical step in the trend towards completely automatic production, which will demand measurement and control of every stage in a manufacturing process.

(2) MATERIALS

Developments in magnetic and synthetic insulating materials have resulted in many improvements in the design and performance of instruments. In some cases, the mere substitution of a better material has been sufficient, but in others, and particularly in magnetic circuits, a complete redesign has been necessary to take full advantage of the enhanced properties of the material.

(2.1) Permanent Magnets

(2.1.1) Dispersion Hardening Alloys.

Dispersion hardening magnets were first discovered by Mishima in 1931, who announced that an alloy of aluminium, nickel and iron showed exceptional magnetic and mechanical hardness.[1] Mishima's work stimulated research in permanent-magnet materials, resulting in the high-energy anisotropic magnets now available. Since 1946 the highest energy content of commercial magnets has increased[2] from 4·8 to 7·5 megagauss-oersteds.

The difficulty of machining dispersion hardening magnets has led to the production of various types of composite magnet (see Fig. 1). The sintered composite magnet is a homogeneous structure of magnet material and soft iron, produced by powder-metallurgy methods, in which the soft iron provides the yokes and pole pieces of the magnetic circuit.

An alternative composite magnet, which has the advantage of using a cast magnet, is produced by casting the magnet directly on to soft-iron pole pieces. The pole pieces may be machined from mild-steel bar or sintered to the desired shape.

In other forms of composite magnet the soft-iron yokes and pole pieces are secured to the magnets either by brazing or bonding with epoxy resins.

(2.1.2) Powdered-Iron Magnets.

Apart from their practical importance, powdered-iron magnets are a development of outstanding scientific interest. It was originally discovered at Grenoble University that, in pure iron powder, in which the particle size is of the order of 10^{-5} cm, coercivities exceeding 400 oersteds are obtained.[3] Commercial methods of producing such micropowders were developed in France.[4] The magnets are produced by pressing the powder to the required shape at room temperature. Complicated shapes can be achieved to a high accuracy and with a good finish. In a further development of these magnets, cobalt is added to the pure iron, resulting in an increase in coercive force at the expense of remanence. The total energy is slightly reduced in the alloy magnet. Micropowder magnets are isotropic, mechanically soft and have energies up to about 1·7 megagauss-oersteds.

(2.1.3) Stability of Magnets.

The extent to which a magnet is affected by an external field depends both on the coercivity and the fullness factor of the magnet.[2] All the magnets considered above have high coercive forces. The hard magnets have fullness factors approaching 0·7, whilst that of the soft magnets varies between 0·3 in the case of the alloy magnets and 0·5 for the pure iron.

It has been shown that artificial ageing is unnecessary with modern magnets.[5] Tests on dispersion-hardening magnets have shown that, during the first three weeks after magnetizing, the strength of magnets varied by approximately 0·3% and thereafter remained constant to within about 0·1%. Some manufacturers, however, demagnetize magnets a few per cent with the dual purpose of adjusting sensitivity and ageing the magnets.

(2.2) Soft Magnetic Materials

Work on soft magnetic materials for instruments has been largely directed towards improving the characteristics of known materials. Notable improvements have been effected both in nickel-iron and in silicon-iron alloys, but these have been achieved by developments in manufacturing techniques rather than by any variations in the alloys. Thus Supermumetal and Special Radiometal are produced by paying extreme attention to the purity of the alloy constituents and by maintaining that purity by melting and annealing at closely controlled temperatures in a high vacuum.[6] The material is rolled down to a thickness usually in the range 0·001–0·010 in, but for particular applications it can be rolled down to 0·00025 in. Special electronically controlled rolling mills are used to ensure accuracy of strip thickness.

(2.2.1) Nickel-Iron Alloys.

Supermumetal and Supermalloy are 77% nickel-iron alloys containing a small proportion of molybdenum and are characterized by extremely high initial and maximum permeabilities and by low loss. The initial permeability, measured with a magnetizing force of 50×10^{-6} oersted, is guaranteed to exceed 5×10^4 with a maximum permeability of 2×10^5. Fig. 2 compares the initial permeability of Supermumetal with standard Mumetal at various frequencies. Fig. 3 shows the losses of the two materials. The alloy requires very careful handling, as the slightest mechanical strain will degrade its magnetic properties, and it is usually supplied in plastic cases which

Mr. Axworthy is with Everett, Edgcumbe and Co., Ltd.

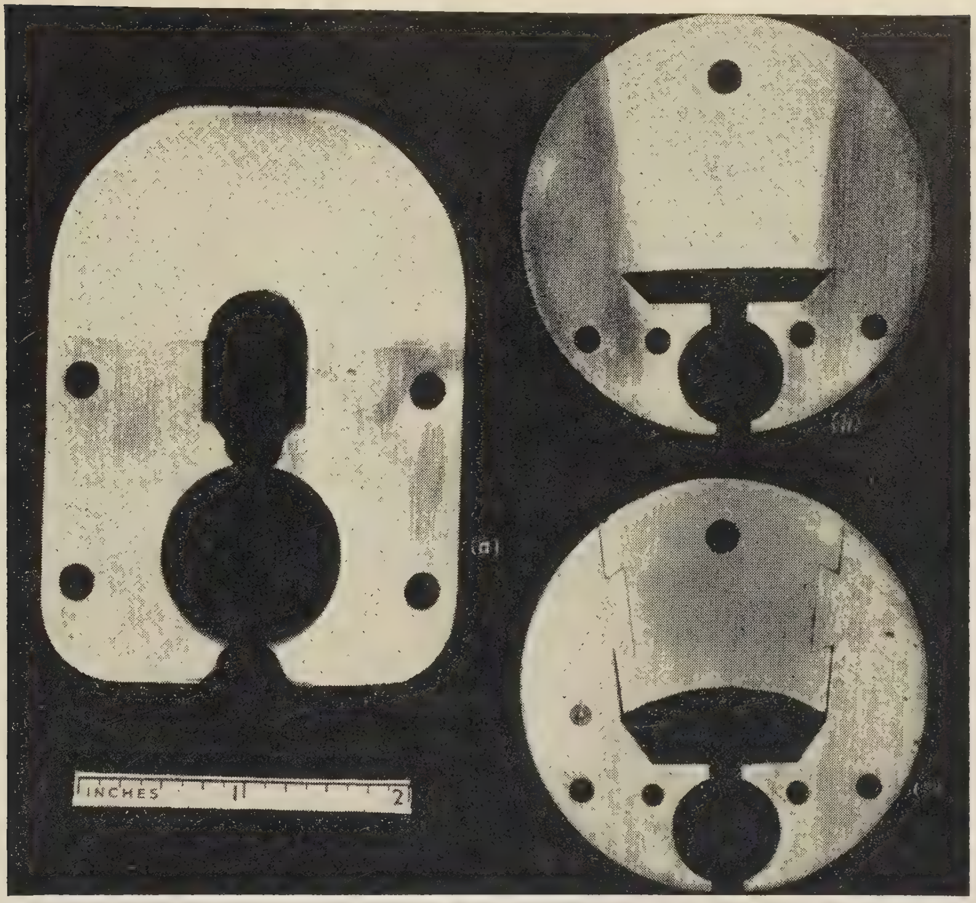


Fig. 1.—Composite magnets.
(*a*) Brazed.
(*b*) Sintered composite.
(*c*) Cast composite.

should not be removed. It can be used for laminations, but again the greatest care, both in handling and in the design of the laminations, is required.

Special Radiometal, a 50% nickel-iron alloy, retains the advantage of the standard grade of having a higher saturation induction than Mumetal. A comparison of the standard and special grades of the material shows that initial permeability is increased from 2 000 to 3 300 and maximum permeability[7] from 25 to 55 × 10³.

A grain-oriented 50% nickel-iron, originally developed in Germany, is now manufactured in Great Britain under the names H.C.R. Alloy or Permalloy F. The grain orientation is obtained by heavily cold rolling the material, which becomes magnetically anisotropic. It has an almost rectangular hysteresis loop and finds its main application in transductors. The initial permeability of H.C.R. is 1 000, the maximum permeability is 50 000 and the saturation induction is 16kG.[8]

The Curie point of nickel-iron alloys varies very considerably with the proportion of nickel. An alloy of approximately 30% nickel, 70% iron, has a Curie point of only 70° C. It is now widely used as a magnetic shunt to provide temperature compensation, particularly of permanent-magnet instruments. Fig. 4 shows how the permeability varies with temperature and magnetizing force.

(2.2.2) Silicon-Iron Alloys.

Cold-rolled 3% silicon-iron alloy is now readily available. Grain orientation, achieved by cold rolling and special annealing techniques, results in a reduction of hysteresis loss to approximately 30% of the loss in a similar hot-rolled alloy, together with an increase in permeability. The initial permeability and total loss of grain-oriented silicon iron are shown in Figs. 2 and 3, respectively. Toroidal cores are widely used for current transformers of all but the highest-accuracy classes, whilst, for voltage transformers, C-cores are displacing laminated-iron circuits.

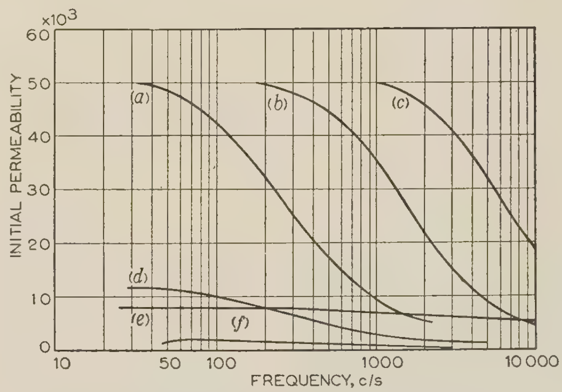


Fig. 2.—Initial permeability of soft magnetic alloys.
(*a*) Supermumetal 0·010 in.
(*b*) Supermumetal 0·004 in.
(*c*) Supermumetal 0·002 in.
(*d*) Mumetal 0·015 in.
(*e*) Mumetal 0·002 in.
(*f*) Grain-oriented silicon iron 0·014 in.

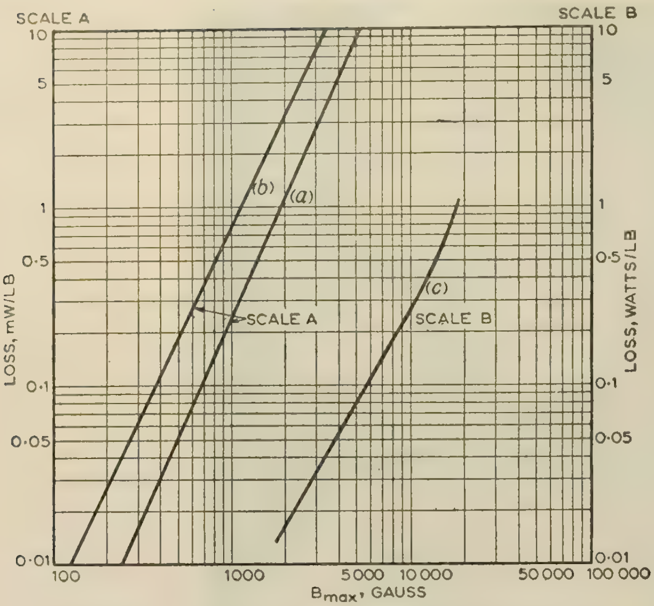


Fig. 3.—Core loss of soft magnetic alloys at 50 c/s.

(a) Supermumetal 0·004 in.
(b) Mumetal precision grade.
(c) Grain-oriented silicon iron 0·014 in.

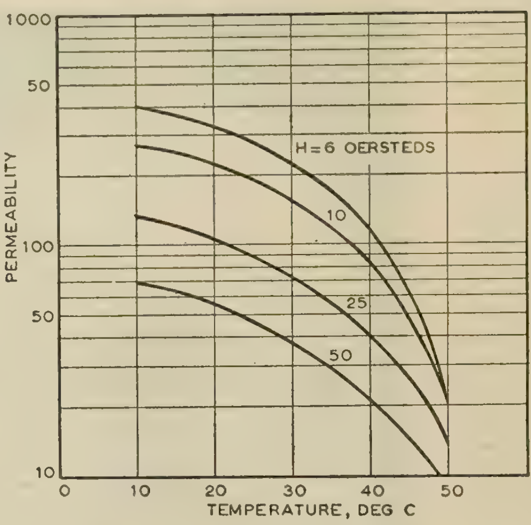


Fig. 4.—Variation of permeability with temperature at various magnetizing forces.

30% nickel-iron alloy.

(2.3) Enamelled Wire

In recent years, there has been intensive development in coverings for winding wires, and particularly in the production of synthetic enamels.[9] Because of its price, oleo-resinous enamelled wire still has a wide application where operating conditions are not severe, but being relatively soft, the covering is easily damaged by abrasion or shock. Of the synthetic enamels, polyvinyl formal is the most widely used, but a number of newer coverings are now available to instrument designers.

The very high abrasion resistance of polyvinyl formal introduces difficulties when soldering fine wires. These are overcome by the use of polyurethene-resin enamel, which has properties similar to polyvinyl-formal films although the abrasion resistance is slightly lower. Polyurethene-covered wires can, however, be soldered at 350° C without removal of the covering.

Polyester-type enamels, derived from Terylene, have been used on the Continent at temperatures up to 150° C, but British manufacturers at present recommend a maximum operating temperature of 130° C for this type of covering. Apart from their high working temperature, polyester enamels are particularly resistant to solvents.

For higher-temperature applications, a silicone-modified alkyd enamel has been developed having a normal maximum temperature of 155° C.

The main properties of enamel coverings are summarized in Table 1.

(2.4) Resistance Wires

The constantan and manganin types of alloy are still the most widely used resistance-wire materials. Newer alloys, whose chief advantages are their higher resistivity and the possibility of modifying their temperature coefficients, are now available.[10, 11] Because of their high specific resistance, the new alloys would seem to be ideal for high-frequency resistors. Their long-term stability has yet to be established, however, and they are unlikely to displace the older materials as instrument resistors until more is known about secular changes of resistance at operating temperatures. Some properties of resistance alloys are shown in Table 2.

Table 1

PROPERTIES OF ENAMELS FOR WIRE COVERING.[9]

Type	Name	Max. service temperature	Flexibility	Resistance to abrasion	Electric strength at 20° C	Electric strength After 24 h immersion in water	Resistance to solvents	Resistance to plastic flow
		°C						
Oleo-resinous	Standard enamel	105	Moderate	Poor	Good	Moderate	Moderate	Moderate
Polyvinyl formal	Lewmex	120	Excellent	Excellent	Good	Poor	Very good	Good
Polyurethane	Lewcosol	120	Very good	Good	Very good	Good	Very good	Good
Polyester	Lewkanex	130	Excellent	Very good	Very good	Good	Excellent	Very good
Epoxy	Lewsyn	120	Good	Moderate	Very good	Good	Good	Good
Modified silicone	Lewsilicon	130–150	Good	Moderate	Very good	Moderate	Very good	Very good
Polyacrilonitrile	—	130	Very good	Very good	Very good	Poor	Excellent	Very good

(2.5) Insulating Materials

Synthetic insulating materials are generally organic polymers, having either long chain molecules in which no permanent chemical bond exists between adjacent molecules and which usually are thermoplastic, or a cross-linked molecular structure

Table 2

PROPERTIES OF RESISTANCE ALLOYS

Alloy	Approximate composition	Resistivity	Thermal e.m.f. to copper
		microhm-cm	μV/deg C
Constantan	60% Cu, 40% Ni	49	40
Manganin	84% Cu, 12% Mn, 4% Ni	44	1
Centanin	67% Cu, 27% Mn, 5% Ni	100	3
Evanohm	75% Ni, 20% Cr 2% Al, 2% Cu	125	1
Karma	75% Ni, 20% Cr 2% Al, 2% Fe	125	1

which is thermosetting. The thermosetting materials are always used with fillers when moulded or laminated.[12,13]

(2.5.1) **Thermoplastic Materials.**

An increasing number of thermoplastic mouldings are being used in instrument manufacture. Choice of the correct material will usually depend on the function of the particular moulding. Some of the more commonly used polymers and their salient properties are listed below. Their more important electrical characteristics are summarized in Table 3.

(*a*) *Polyethylene.*—A flexible but very tough plastic with excellent insulating qualities. It has a maximum working temperature of about 70° C.

(*b*) *Polystyrene.*—A brittle material having low dielectric loss. It is chemically stable, easily moulded and may be transparent or opaque. It has exceptionally low thermal conductivity and a maximum working temperature of 70° C.

(*c*) *Polymethyl methacrylate.*—Tougher than polystyrene but equally rigid, Perspex has poor dielectric qualities although the surface resistivity is very high. It is very transparent and is sometimes used as a substitute for glass. It can be used up to about 70° C.

(*d*) *Polyamides.*—Owing to its hygroscopic nature, nylon is not a particularly good insulator. On the other hand, it has exceptional strength. It has a low coefficient of friction and a relatively high softening temperature, being usable up to 100° C.

(*e*) *P.T.F.E.*—Unlike the materials mentioned above, p.t.f.e. cannot be moulded. It is sometimes extruded, but usually powder is compressed to the required shape and sintered. P.T.F.E. is remarkable in many respects. Chemically, it is entirely inert, and being also strongly hydrophobic, it maintains its high surface resistivity in all conditions. It is also noteworthy in that its static and kinetic coefficients of friction are equal. Use of the material in industrial-instrument design is restricted to some extent by its high cost.

(2.5.2) **Thermosetting Plastics.**

An important development in thermosetting plastics is the increasing use of epoxy resins. Apart from their use as a potting medium, the epoxide materials adhere with great strength to metals, plastics, wood, glass and most other solids, providing a very satisfactory method of bonding components. Thus all the inactive metal parts of the moving element of an instrument can be made of Duralumin or aluminium secured with an epoxy adhesive, resulting in very considerable savings in weight. Epoxy resins are also used as protective coatings, which are highly resistant to attack by corrosive atmospheres, have high abrasion resistance and satisfactory insulating qualities.

(2.6) Silicones

The use of silicone fluids and greases and silicone-modified materials in instrument manufacture is increasing rapidly. Their most useful properties are their wide temperature range, −50° C to 150° C, their strongly hydrophobic nature, and, with regard to fluids, their almost constant viscosity.[14] Silicone damping fluids and silicone rubber gaskets for sealing instrument cases are obvious applications.

Table 3

ELECTRICAL CHARACTERISTICS OF THERMOPLASTICS

	Polythene	Polystyrene	P.T.F.E.	P.V.C.	Nylon	Perspex
Maximum working temperature, °C	70	70	250	70	100	70
Thermal conductivity, c.g.s. units	7	1·9	6	3·5	6	3·5
Surface resistivity, ohms	10^{14}	10^{13}	10^{13}	10^{12}	—	10^{14}
Volume resistivity, ohm-cm	10^{17}	10^{13}	10^{17}	10^{15}	10^{14}	10^{13}
Power factor at 10 Mc/s	0·0002	0·0004	0·0001	0·02	0·02	0·02
Dielectric constant at 10 Mc/s	2·3	2·6	2·0	3·2	4·0	3·2
Breakdown voltage, volts/mil	1000	400	400	375	300	390

(3) DESIGN DETAILS

Changes in design are mainly due to the application of new techniques, new materials and changes in users' requirements. In some cases, e.g. the use of printed circuits, techniques have been acquired from other branches of the industry, whilst, in others, special techniques have had to be evolved. Hermetically-sealed instruments, which resulted from a specific demand by users, provide an example of the latter.

(3.1) Printed Circuits

Originally developed by the electronic industry, printed circuits are now being applied to instrument manufacture. A new battery-driven insulation tester, having a self-contained transistor-operated voltage convertor, uses printed wiring throughout with conventional components. Conversely, a recently

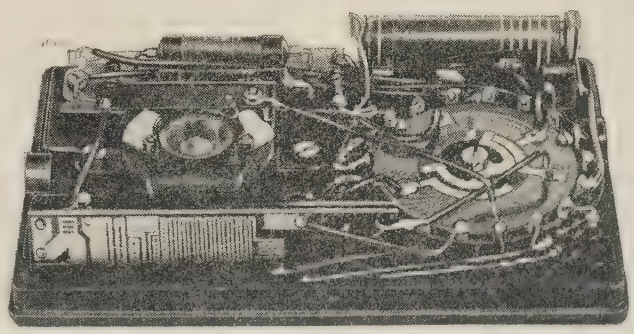


Fig. 5.—Multi-range test set with printed resistors.

introduced multi-range test set, shown in Fig. 5, which is wired normally, is probably the first electrical indicating instrument to incorporate printed resistors. These are produced from a special alloy. One of the printed resistance cards is designed as an auxiliary switch plate to form an integral part of the selector switch.

slot of a coil. The moving iron is in the form of part of a hollow cylinder whose axis coincides with the axis of deflection.[17] This instrument is extremely efficient, the full-scale torque being 0·6 g-cm for an input of 0·8 VA. Eddy-current damping is used.

Two types of wattmeter are in common use for long-scale instruments. Several new induction movements have been designed, which, with their naturally long scale, are ideal for normal power-frequency applications. Iron-cored dynamometers are also produced, having a magnetic circuit which is similar to that of a radial-field permanent-magnet instrument. For use as a wattmeter, the magnet is replaced by a wound coil. Neither is particularly suitable for use outside the power-frequency range, but an interesting possibility for a long-scale audio-frequency wattmeter is provided by the crystal wattmeter described in a recent paper.[18] This utilizes the Hall effect of germanium or indium-antimonide crystals—a permanent-magnet moving-coil instrument being used as the indicator.

In the smaller sizes of long-scale frequency meter, the most commonly used arrangement is a bridge network feeding a rectifier moving-coil instrument. Such bridges can be used either for power-frequency or audio-frequency applications.

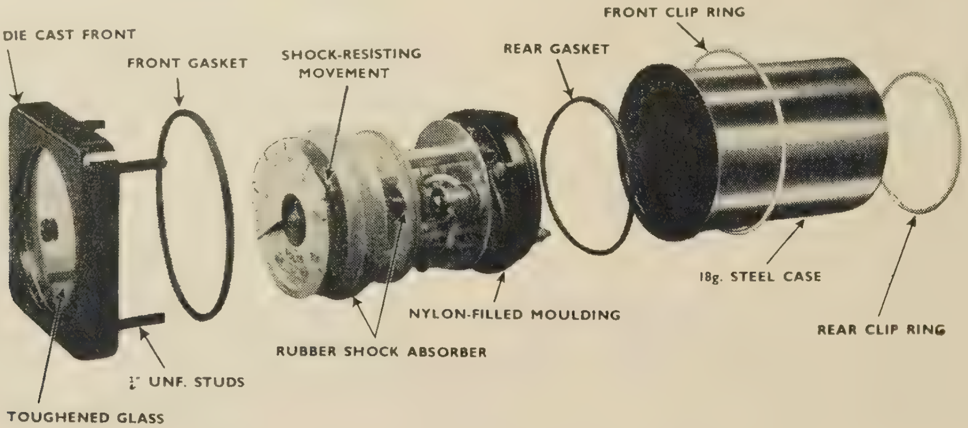


Fig. 6.—Hermetically sealed shockproof instrument.

(3.2) Sealed Instruments

Hermetically-sealed and shockproof instruments were originally derived from Services' requirements. The Service pattern, which is also used for industrial applications, has a 3½ in diameter circular dial giving a scale length of over 7 in. Mounted in pressed-steel cases, they are sealed at either end by a silicone-rubber sealing ring. The zero adjuster is also sealed with a silicone-rubber gasket. After assembly, the cases are evacuated, filled with dry nitrogen and sealed off. An 'exploded' view of such a sealed and shockproof instrument is shown in Fig. 6. Other types of sealed instrument, in sizes ranging from 1 to 8 in-diameter dials, have silicon-aluminium die-cast cases which are sprayed with an anti-fungus enamel. The specification for such instruments[15] requires that they should withstand a temperature range of $-40°$ C to $+70°$ C, 95% relative humidity and a pressure differential of 10 lb/in^2.

(3.3) Long-Scale Instruments

A feature of electrical indicating instruments since the war has been the increasing use of movements whose pointers subtend an angle between 240° and 260°. A number of completely new movements have been designed for these instruments.

In the permanent-magnet versions, the coil may work either with a parallel magnetic field co-operating with the coil sides normal to the axis or in a radial field which reacts with the side of the coil remote from and parallel with the axis. The latter arrangement allows a much more efficient magnetic circuit than the first, and some designs of radial-field instruments approach conventional short-scale types in their overall efficiency. Such a movement,[16] illustrated in Fig. 7, has a full-scale torque of 1·1 g-cm for an input of 3·5 mW.

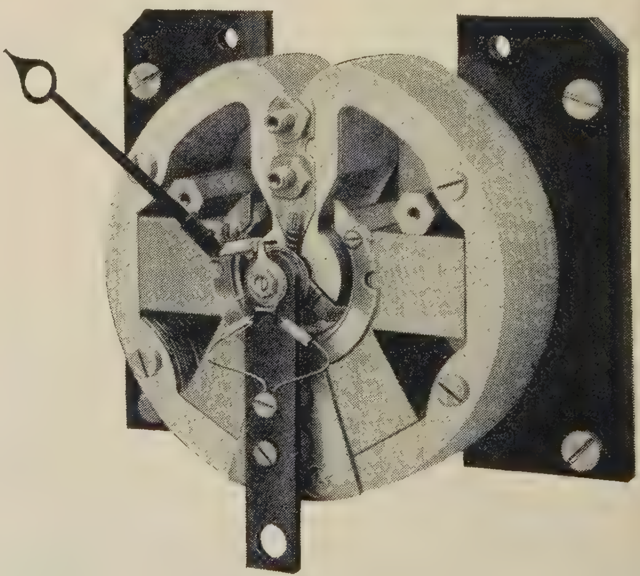


Fig. 7.—Long-scale permanent-magnet moving-coil instrument.

Long-scale moving-iron instruments usually have combined repulsion-attraction movements. Over the first half of the scale the moving iron is repelled by a fixed iron and is then attracted towards the gap between two further fixed irons over the rest of the scale. An overlap between the two sections ensures that no instability occurs at the change-over point. A typical instrument of this design has a full-scale torque of 0·36 g-cm for an input of 1·75 VA. An alternative design uses an attraction movement, with a tapered moving iron passing through the

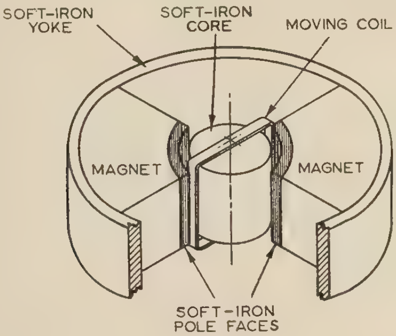
The series is completed by power-factor indicators which may be either moving-iron instruments with a 360° scale or dynamometers in which the scale angle is restricted to about 270°.

All the above movements have been designed for use with $\frac{1}{2}$-in-diameter scales, and most of the leading British manufacturers now produce a complete range of such instruments.

(3.4) **Permanent-Magnet Instruments**

In conventional permanent-magnet moving-coil instruments of 90–120° scale angle, composite magnets are now becoming widely used in place of fabricated designs. A further trend has been toward the design of improved magnetic circuits.

One type of movement, shown in Fig. 8(*a*), has a magnet



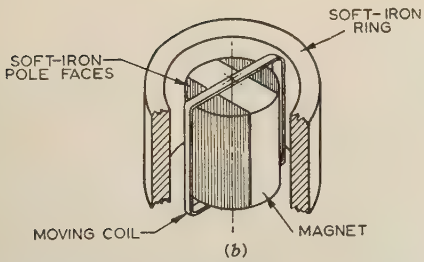


Fig. 8.—Magnetic circuits of short-scale moving-coil movements.
(*a*) External magnets.
(*b*) Internal magnets.

adjacent to each pole face with the circuit completed by an iron yoke surrounding the whole assembly. The magnet system, which is entirely self-shielding, has a relatively low leakage coefficient. In one version, the magnets and pole faces form composite sintered units, whilst in an alternative design a fabricated assembly with cast magnets is preferred. Because of the inherently efficient magnetic circuit, high gap fluxes are achieved with small magnets.

The internal magnet movement was made practicable by the production of high-energy magnets and was first introduced before the war.[15] Its use was, however, restricted to one manufacturer until the advent of sintered composite magnets. The composite magnets, cylindrical in shape and diametrically magnetized, replace the soft-iron core of a conventional instrument, whilst a soft-iron annular ring is substituted for the normal yokes [see Fig. 8(*b*)]. The circuit is self-shielding and of extremely high magnetic efficiency; for example, a conventional magnetic circuit with a magnet of $36\,cm^3$ volume can be replaced by an internal magnet system with $6\cdot5\,cm^3$ of magnet, the two arrangements giving the same effective gap flux density. A peculiarity of internal magnet systems, if correctly designed, is that the flux density calculated from the constants of the instrument is some [illegible]% higher than the measured flux density. This indicates that a substantial torque is produced by the top and bottom as well as by the sides of the coil moving in the air-gap. A disadvantage of core-magnet instruments is their restricted scale angle—90° is about the maximum deflection for an evenly-divided scale compared with 120° for more conventional types.

(3.5) **Moving-Iron Movements**

Recent designs of moving-iron instruments do not depart seriously from established practice. A wide extension of the frequency range of a moving-iron attraction-pattern voltmeter was achieved by significantly increasing the sensitivity of the movement. Increased sensitivity was attained by the addition of iron to the coil. The errors of such an instrument do not exceed 1% between 50 and 2 500 c/s with sinusoidal, triangular or square waveforms. A recently published analysis of the shunt-capacitor method of frequency compensation permits a precise calculation of the capacitance required and the range of frequency over which it is effective.[20] A better understanding of factors controlling scale shape, coupled with improvements in manufacturing techniques, has allowed the use of pre-printed scales on repulsion ammeters and voltmeters of industrial accuracy. Scale shapes can be repeated on quantity-produced instruments to within $\pm 0\cdot75\%$ of a standard scale. Of particular interest is a moving-iron movement of the parallel-iron type, which has almost ideally linear scale from 5% up to 100% of the scale range. This result is obtained by siting the axis of the movement to one side of an elliptical coil rather than in the centre of a round one.[21]

(3.6) **Wattmeter Movements**

Induction movements and air-cored dynamometers are the most commonly used types of wattmeter movement for panel instruments, although the thermocouple wattmeter finds some applications, mainly for audio-frequency work and in aircraft.[22]

The use of oval, in place of circular, coils, in iron-less dynamometer movements, gives a useful saving in the overall depth of the movement. In a typical design, the ratio of major to minor axis of the current coil is $1\cdot5 : 1$. This technique is particularly effective when applied to polyphase instruments in which two or three elements are mounted in tandem on a common spindle.

(3.7) **Presentation**

Since the war, the importance of applying scientific methods to the presentation of measured data has been increasingly recognized. Considerable experimental work on the legibility of numbers and the readability of various forms of scaling has been carried out, particularly at Cambridge University by the Applied Psychology Unit of the Medical Research Council.[23] At the same time, attention has been paid to the overall appearance of instruments by the introduction of qualified industrial designers responsible for the styling of instruments.

(3.7.1) **Scales.**

It has been established that the simpler the graduation system consistent with supplying the required information, the more easily will the instrument be read.[24] The modern tendency is to reduce irrelevant lettering on the scale plate to a minimum. Very often, the quantity being measured, the maker's name or trade mark, the accuracy grade and the instrument number, the last two in small lettering at the bottom of the dial, are all that appear, apart from the graduations. Any other necessary data are shown on the back of the instrument. Unfortunately, some Service specifications still require that the scale plate be 'cluttered up' with writing.

The combination of knife-edge pointer and mirror scale is still thought to provide the most accurate method of reading an

instrument and is universally used on precision-grade instruments. The platform scale, in which the graduated part of the dial is raised to the plane of the pointer, is often fitted as an anti-parallax device.[15, 25] In a particular series of long-scale instruments, the platform scale is made of Perspex, into which two small lamps are inserted at the corners. These lamps, concealed from the front, provide an annular ring of light around the graduated part of the scale.

During the period under review, there has been an increasing demand for flush rectangular instruments in size up to about 6 in square. Some manufacturers have met this demand with completely rectangular cases, but round-bodied instruments having rectangular fronts are now usually preferred. Moulded cases are generally acceptable, and where these cannot be used, diecast aluminium is often substituted for instruments up to about 5 in square. In the larger sizes, pressed steel is ousting cast iron.

(4) MEASUREMENT TECHNIQUES

(4.1) A.C. Bridge Networks

Bridge networks are being used in a growing number of instruments for certain a.c. measurements. A number of manufacturers are now producing frequency meters in which a permanent-magnet moving-coil indicator shows the degree of balance of a frequency bridge network. In one example, which can be used in the power-frequency band and in frequencies up to 20 kc/s, two double-T bridges are used, one balanced above and one below the range of the instrument. The current consumption of such an instrument is of the order of 3 mA. Ring-modulator circuits are used for measuring the component of current in a circuit which is in phase with the voltage.[26] An interesting application of such an instrument is the wattmeter bridge, designed for production testing, in which the indicator is scaled in percentage deviation from a predetermined value.

Other instruments incorporating bridge networks include synchroscopes and phase-sequence indicators.

(4.2) Speed Measurement

Except where direction of rotation is involved, a.c. tachometer generators with rectifier-operated indicators are almost universally used for speed measurements, although some makers prefer to use a d.c. generator for low speeds. Three-phase alternators, used in conjunction with full-wave 3-phase bridge rectifiers and permanent-magnet moving-coil indicators have recently been introduced. A 12-pole 3-phase generator without gearing is used for indicating speeds as low as 50 r.p.m. (full scale), while at the other end of the range, speed indicators having a range of 0–80 000 r.p.m. are available. The latter instruments measure the speed of cold-air turbines in aircraft. The 2-pole magnet is bolted directly on to the turbine rotor shaft and the stator clamped to the body of the turbine. To provide for interchangeability of generators and indicators, rectifier-operated ratiometers are used to measure the output frequency.

(4.3) Resistance Measurements

Ratiometer instruments are still the most widely used for the measurement of resistance. A long-scale version, which is also applied to the measurement of temperature and frequency, is illustrated in Fig. 9, which shows an 'exploded' view of the movement together with a view of the instrument arranged as a tap-position indicator.

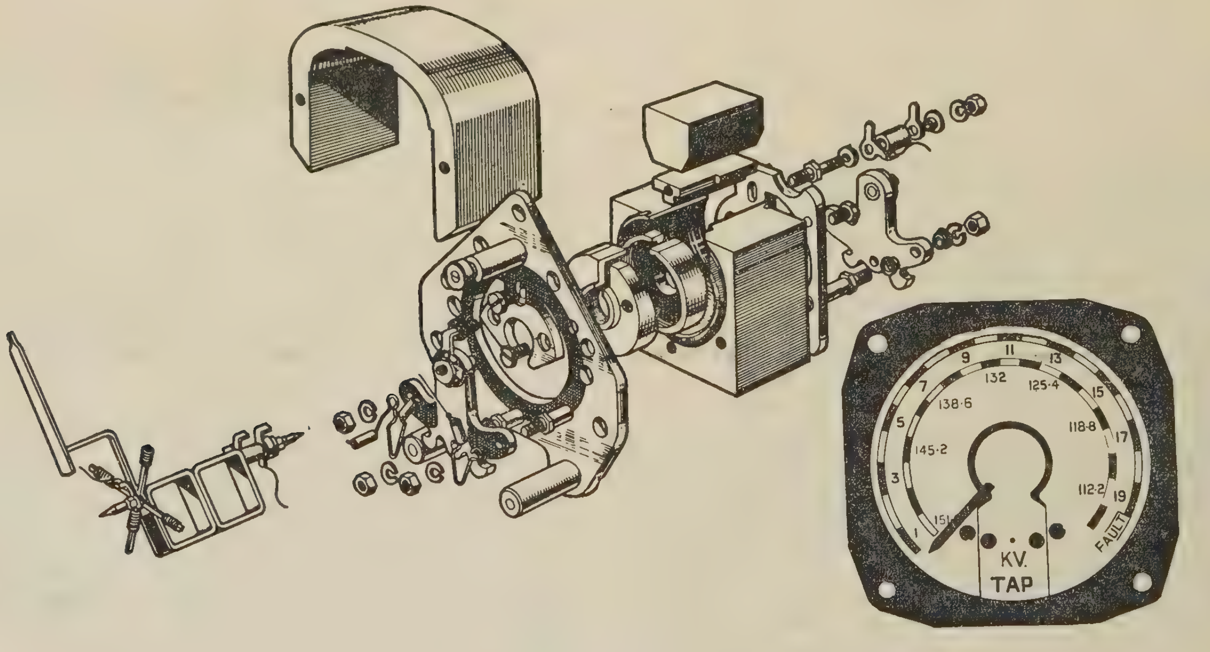


Fig. 9.—Long-scale ratiometer movement.

Milliammeter methods of measuring resistance are increasing in popularity. An unconventional instrument, designed for production testing, comprises a stabilizer feeding a Wheatstone bridge which has a valve voltmeter, calibrated in terms of percentage deviation from the present desired value, in place of the usual galvanometer. A valve voltmeter is also used as the indicator of an ohmmeter, which is noteworthy in having an evenly-divded ohms scale. The instrument, which has full-scale readings ranging from 3 ohms to 1 megohm in 12 steps, feeds a stabilized current into the unknown resistance. The voltage across the resistance is then measured by the valve voltmeter.

Recent developments in the design of insulation-resistance

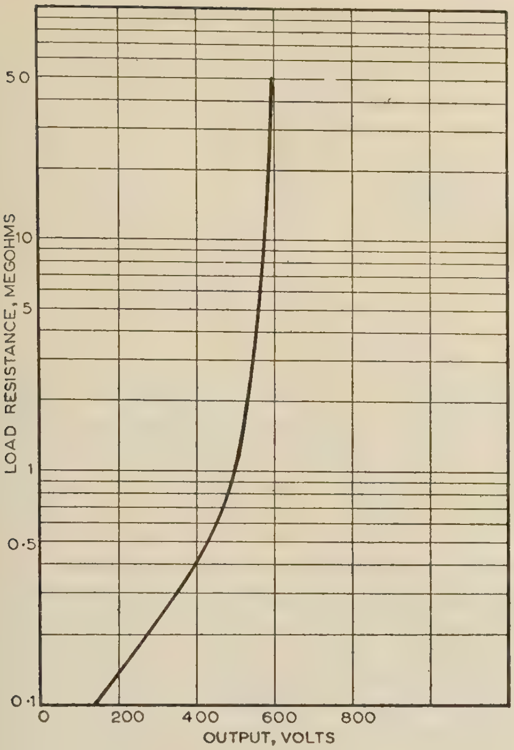


Fig. 10.—Regulation of transistor d.c. convertor.

testers include the incorporation of a transistor-operated voltage convertor to provide a high d.c. test voltage from a small 9-volt battery. The higher efficiency obtainable with transistor convertors ensures a long life from the battery, whilst its ease of operation is a particularly attractive feature of the instrument. The regulation curve of the instrument is shown in Fig. 10.

The 13th Edition of The Institution's Regulations for the Electrical Equipment of Buildings requires a measurement of the earth-fault loop impedance by a current-injection method, and several instruments have been designed to measure the impedance of the neutral-earth loop with injected currents up to 20 amp. Most instruments measure current at a fixed voltage with the ammeter also having a resistance scale. Owing to the voltage which may exist between earth and neutral, the makers usually recommend that two tests be made at opposite polarities and the mean of the two resistance readings be taken. Tagg[27] has shown that considerable inaccuracies may result if a phase displacement exists between current in the neutral and the injected current, and G. L. d'Ombrain pointed out, in the discussion on Tagg's paper, that a more logical method would be to calculate the impedance from the mean of the two current readings. An alternative design uses a bridge network with the neutral-earth loop impedance forming one arm of the bridge, again requiring two readings to be made at opposite polarities. A d.c. instrument has been designed in which a hand-driven generator provides the injected current, which is continually reversed by a commutator. The particular Regulation concerned, Regulation 507 Section 5, has received some adverse criticism, and other methods of measuring the resistance of the earth-fault loop have been proposed.[28]

(4.4) Power Measurements

Measurements of active, apparent and reactive power and power factor are made for monitoring power supplies. Any two of these are sufficient for a complete appreciation of the load

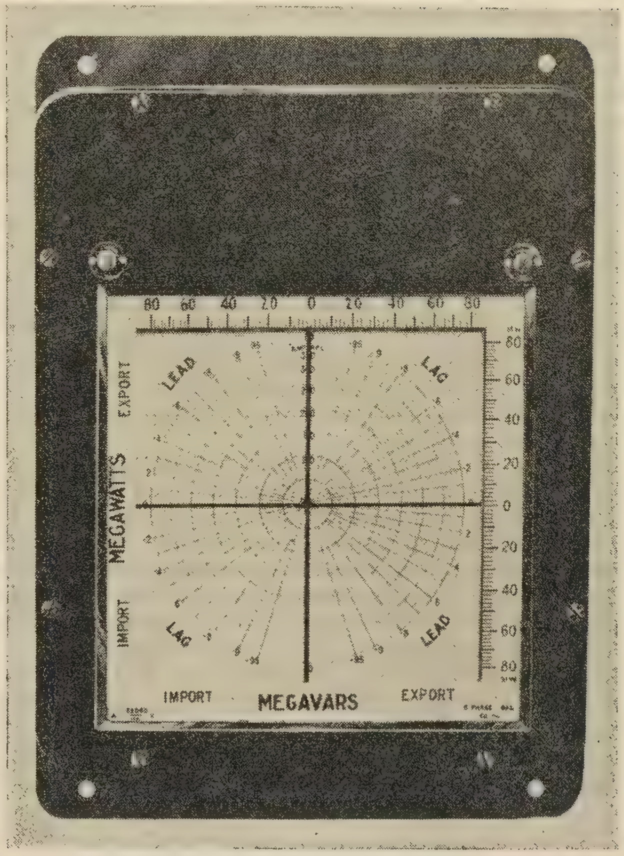


Fig. 11.—Vectormeter.

conditions, and a measurement of active power is almost invariably included. The Vectormeter (Fig. 11), however, provides all the relevant information in a single instrument.[15] Two pointers moving normally to each other and in a plane parallel to the scale, measure active and reactive power, respectively, and the intersection of the pointers give both current and power factor. Such instruments are mainly used in power-station control rooms.

An alternative is the combined wattmeter and varmeter in which a single instrument measures active and reactive power successively, the quantity being selected by a switch. Only one voltage transformer and one current transformer are required for both measurements.

Direct-reading volt-ampere meters are achieved by feeding the coils of a dynamometer movement from 3-phase rectifier networks.

An interesting proposal for measuring power factor uses a small synchro as the movement. The 3-phase winding of the machine has its star point opened and the coils connected to a 3-phase supply. The other ends of the coil are then connected to a second supply in such a manner that the individual supplies produce counter-rotating fields. The H-type rotor of the synchro then aligns itself in a position determined by the phase difference of the two supplies.[29] It is claimed that the method produces a power-factor meter of phenomenally high torque.

(4.5) Telemeters

Telemeters find their main application in the remote indication of summation-wattmeter readings in power stations, of pressure-gauge readings, of gas and water flow and of liquid levels. Most repeaters operate on the torque-balance principle and the

majority include electronic elements. A notable exception, originally due to G. F. Shotter, which uses two iron circuits with a variable coupling as the transmitter, has recently been redesigned to take advantage of modern magnetic materials and techniques. A widely used electronic repeater is based on a well-known photocell-operated d.c. amplifier.[30] When the amplifier is modified to operate as a telemeter, the originating element takes the place of the usual galvanometer and is mechanically controlled by a permanent-magnet moving-coil instrument which replaces the feedback resistance and which mechanically applies negative feedback in the form of a control torque on the originating spindle.[31] The Marchment electronic remote indicator uses a hard valve, the grid bias of which is controlled by contacts operated by a moving-coil relay in the anode circuit. The average current through the relay produces a force which continuously and precisely balances the torque of the originating movement.[32]

The output circuit of remote-indicator equipments usually contains a number of moving-coil instruments and often includes indicators up to 48 in square. Instruments of 20 in diameter and larger usually have geared movements, although one manufacturer uses a long-scale radial field movement for all sizes up to 30 in. Selsyn units, operated by a local servo-mechanism responding to the repeater output, are often preferred to geared moving-coil movements in the largest sizes of instrument.

Where the output is to be transmitted over Post Office lines, barrier transformers are required at both ends. In such cases, saturable reactors are used to convert the d.c. output to alternating current and at the same time to amplify the output to provide for the losses in the transformers and transmission lines.

(4.6) Self-Balancing Instruments

Originally developed as recording instruments, electronic self-balancing indicators combine extreme sensitivity with a high degree of ruggedness. They are ideal for measuring low d.c. potentials and are being increasingly applied to the indication of temperature, CO, CO_2, pH, etc. The measuring circuit is either a Wheatstone bridge or a potentiometer, having a d.c. amplifier, usually of the chopper type, as a detector. The output of the amplifier feeds a servo motor which drives a slide wire to maintain balance. The instrument pointer is coupled to the slide wire. A typical instrument has a minimum full-scale range of 1 mV with an error of 0·5%.

A different type of self-balancing instrument provides a d.c. output in milliamperes proportional to a millivolt input, an input of 20 mV giving a current of 20 mA into 1 000 ohms. The instrument consists of a balanced resistance bridge network, one arm of which is an electronic oscillator behaving as a variable d.c. resistance. The oscillator is of the tuned-anode tuned-grid type working under class-C conditions. When the resonant grid circuit is damped, the valve bias becomes more positive and decreases the effective resistance of the valve. The effective resistance is controlled by a permanent-magnet moving-coil indicator in which the pointer is replaced by a vane and the scale by a small pancake coil. The coil is part of the oscillator grid circuit and controls the resistance of one arm of the bridge. A milliammeter, reading the output current, is connected in series with a feedback resistor across the bridge. The moving coil of the control instrument, which is known as an 'Inductrol', is connected in series with a millivolt input across the feedback resistor. When a potential is applied to the input terminals, the Inductrol vane approaches the pancake coil and causes a current to flow across the bridge. The current increases until the voltage drop across the resistor balances the input potential. The current through the feedback resistor, which also flows in the output circuit, is directly proportional to the input potential. A simplified circuit is shown in Fig. 12. A feature of the instrument is its small size, the unit occupying only $3\frac{1}{2}$ in × $7\frac{1}{4}$ in × $8\frac{1}{2}$ in.

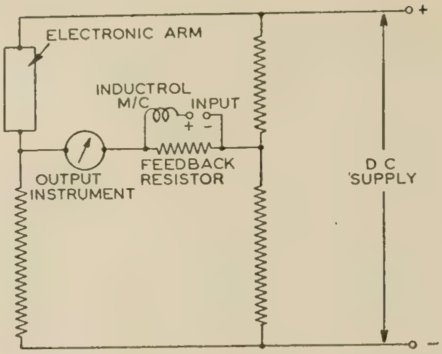


Fig. 12.—Simplified circuit diagram of the servo-potentiometer.

(4.7) Audio-Frequency Power Measurement

The use of specially designed feedback amplifiers in conjunction with dynamometer wattmeters has provided a considerable extension both of the frequency range and the sensitivity of the instruments.[33] An alternative approach recognizes the inherent virtues of the electrostatic wattmeter for the measurement of power at high frequencies. A recent paper discussed the use of the electrostatic instrument for iron-loss measurements.[34] It was shown that the usual difficulties of setting up the instrument were removed by arranging that a small section of one quadrant was adjustable in a vertical direction. It was also demonstrated that, if amplifiers be used, the error due to the quadrant shunt will be eliminated if the gain of the amplifier in the current circuit is twice that in the voltage circuit. The paper resulted in the production of a portable electrostatic wattmeter[35] which is particularly suited to the measurement of iron loss at frequencies up to 30 kc/s.

Although the principle on which thermocouple wattmeters operate was established in the early part of the century, it is only recently that they have become available as production-model portable instruments for audio-frequency measurements. Six-inch-scale wattmeters are now made for use up to 5 kc/s with ranges between 2·5 and 600 watts full scale and a maximum error of ±3%.

(4.8) Frequency Monitoring

The pendulum-controlled master frequency meter continues as a key instrument in control rooms. A recent development, in which a crystal-controlled oscillator replaces the pendulum as a time standard, uses a digital display showing time error in minutes and seconds fast or slow.

The use in control rooms of instruments indicating the rate of change of frequency is increasing. These instruments, complementary to the usual precision frequency meters, help control engineers to anticipate load changes.

An electronic instrument obtains the rate of change by comparing the mains frequency with the output of a crystal oscillator. By charging two condensers in turn from a direct voltage proportional to the difference frequency, and comparing the charges on the condensers at 1 min intervals, the mean rate of change of frequency is given.[36]

Another instrument uses two synchronous motors, one driven from the mains supply and the other from a standard source. The difference in the rotation of the two motors gives the time error and is passed to two mechanical differentiators in series,

the output of the first indicating frequency and that of the second the rate of change of frequency.[37]

A third method incorporates a self-balancing frequency meter. The movement of the balancing motor shaft is transmitted to centre-tapped variable potentiometers. Six potentiometers are used, the motor being coupled to each in turn for a 2 min period. Every $\frac{1}{3}$ min a potentiometer is uncoupled and replaced by another. The voltage on the uncoupled resistor, proportional to the mean rate of change of frequency during the preceding 2 min, is measured, after which the slider is returned to the centre position and eventually recoupled to the motor.[38]

(4.9) Live-Conductor Testers

When maintenance work is necessary on high-voltage lines or equipment it is desirable to ensure that the high-voltage conductors are dead before earthing them. The E.R.A. examined this problem and recommended a high resistance in series with a rectifier moving-coil indicator as the most satisfactory means of making the necessary tests.[39] An instrument based on these recommendations comprises a long insulated tube having a contact piece at one end and a $3\frac{1}{4}$ in instrument at the other. A handle is fitted below the instrument with an earthing lead attached to the end of the handle. The insulated tube contains a string of high-stability carbon resistors, the tube length varying according to the maximum voltage on which it is to be used. An instrument for use on 33 kV systems is scaled 0–40 kV, the resistance tube has a length of 30 in and the overall length of the stick is 56 in. Arrangements are made enabling the sticks to be used for phasing-out, whilst bent-end adaptors are available for inserting the sticks into circuit-breaker spouts. A small test box is provided for determining, before and after use, that the instrument is functioning correctly.

Electrostatic live-line testers are similar in many respects to the series-resistance model except that the indicator is an electrostatic voltmeter and the series impedance is a condenser. An extension of the high-voltage terminal into the insulated tube forms one electrode of the condenser whilst an outer metal sheath over the tube provides the other plate. The range of the instrument can be varied by sliding the outer metal tube to vary the coupling. Two ranges, in the ratio of 1 : 2, are usual. Adaptors and testing arrangements are provided.

(5) PORTABLE INSTRUMENTS

The use of long-scale indicators in industrial-grade portable instruments is increasing. The extra length of scale is attractive and makes for easier readability. Some makers fit mirror scales on long-scale portable instruments, although the accuracy of measurement hardly warrants such a refinement.

Multi-range a.c./d.c. instruments using metal rectifiers for a.c. measurements are now firmly established as general-purpose test sets. A new model having 19 ranges uses printed resistors and measures only $5\frac{5}{8}$ in $\times$ $3\frac{5}{8}$ in $\times$ $1\frac{5}{8}$ in. A somewhat large instrument, having a 5 in scale, provides 18 ranges with pushbutton range selection.

Improvements in precision instruments are largely due to the improved materials used in their construction. For example, by using the new nickel-iron alloys in moving-iron instruments, the difference between a.c. and d.c. calibration is reduced to 0·05%, which approaches the limit of reading sensitivity for such an instrument. Better reading sensitivity and accuracy are achieved by the Hawkes–Shotter comparator, which is now widely used as a commercial standard.[40] The instrument can be used for power, voltage and current measurements with a reading sensitivity of 0·01% and a maximum error of 0·05%.

(6) ACKNOWLEDGMENTS

The co-operation of the many manufacturers who supplied information and illustrations regarding their products greatly simplified the preparation of this review, and the author is grateful to the following firms: Avo, Ltd., Bakelite, Ltd., The Cambridge Instrument Co., Ltd., Clare Instrument Co., Crompton Parkinson, Ltd., Electrical Instrument Co. (Hillington), Ltd., Elliott Bros. (London), Ltd., Everett, Edgcumbe and Co. Ltd., Evershed and Vignoles, Ltd., Ferranti, Ltd., The London Electric Wire Co. and Smiths, Ltd., Metropolitan-Vickers Electrical Co., Ltd., Mullard Electronic Products, Ltd., Murex, Ltd., Nalder Brothers and Thompson, Ltd., The Record Electrical Co., Ltd., Sangamo Weston, Ltd., and Telcon Metals, Ltd.

(7) BIBLIOGRAPHY

(1) SALPETER, J. L.: 'Permanent Magnets', *Journal of the British Institution of Radio Engineers*, 1948, **8**, p. 211.

(2) TYRRELL, A. J.: 'The Performance and Stability of Permanent Magnets', *ibid.*, 1950, **10**, p. 812.

(3) BRAILSFORD, F.: 'Ferromagnetism in Relation to Engineering Magnetic Materials', *Proceedings I.E.E.*, Paper No. 1947, February, 1956 (**103** A, p. 39).

(4) British Patents Nos. 596875 and 596966. 1947.

(5) KNIGHT, S. F.: 'A Sensitive Balance for Stability Tests on Permanent Magnets', *Proceedings I.E.E.*, Paper No. 850 M, August, 1949 (**96**, Part II, p. 635).

(6) RANDALL, W. F., and SCHOLEFIELD, H. H.: 'Some Metallurgical and Structural Factors affecting the Properties of Soft Magnetic Materials', *ibid.*, Paper No. 914 M, November, 1949 (**97**, Part II, p. 133).

(7) SOWTER, G. A. V.: 'Some Special Characteristics of Soft Magnetic Materials used in Instrument Manufacture', *ibid.*, Paper No. 1154 M, May, 1951 (**98**, Part II, p. 714).

(8) BARDELL, P. R.: 'Magnetic Materials in the Electrical Industry' (Macdonald, 1955).

(9) TAYLOR, A. J.: 'Winding Wires', *Electrical Manufacturer*, July, 1957, p. 8.

(10) PETERSON, C.: 'Alloys for Precision Resistors', *Proceedings of the N.P.L. Symposium on Precision Electrical Measurement*, 1954, Paper No. 14.

(11) ARNOLD, A. H. M.: 'Alternating-Current Resistance Standards', *Proceedings I.E.E.*, Paper No. 1512, June, 1953 (**100**, Part II, p. 319).

(12) GARTON, C. G.: 'The Properties of Insulating Materials used in Instruments', *ibid.*, Paper No. 1150, May, 1951 (**98**, Part II, p. 728).

(13) COUZENS, E. G., and YARSLEY, V. E.: 'Plastics in the Service of Man' (Penguin Books, Ltd., 1956).

(14) HAYWARD, B. F. W.: 'Silicone Insulation', *The Electrical Manufacturer*, November, 1957, p. 6.

(15) LIPMAN, C. L.: 'Developments in the Design of Electrical Indicating and Recording Instruments', *Beama Journal*, November, 1955, p. 172.

(16) British Patent No. 657005. 1949.

(17) British Patent No. 723938. 1952.

(18) BARLOW, H. E. M.: 'The Design of Semi-Conductor Wattmeters for Power-Frequency and Audio-Frequency Applications', *Proceedings I.E.E.*, Paper No. 1778 M, November, 1954 (**102** B, p. 186).

(19) TAGG, G. F.: 'Graphic Aids to Permanent Magnet Design for Moving Coil Instruments', *Instrument Practice*, October, 1952, p. 828.

(20) MILLER, J. H.: 'Frequency Compensation of Instruments', *Electrical Engineering*, June, 1951, p. 494.
(21) PAULER, J. H.: 'The New Moving Iron Precision Instrument', *Siemens Zeitschrift*, 1952, **26**, p. 129. E.R.A. Translation 1B. 1264.
(22) HAWKES, H. D., and ARNOLD, A. H. M.: 'Wattmeters', *Journal I.E.E.*, 1955, **1** (New Series), p. 676.
(23) BARTLETT, SIR F., and MACKWORTH, N. H.: 'Planned Seeing' (H.M. Stationery Office, 1950).
(24) MURRELL, K. F. H.: 'The Visual Presentation of Instrument Data', *Transactions of the Society of Instrument Technology*, 1952, **4**, p. 1.
(25) GOLDS, L. B. S.: 'A Survey of Modern Methods of Presentation of Instrument Readings and Recording', *Proceedings I.E.E.*, Paper No. 1152, May, 1951 (**98**, Part II, p. 671).
(26) DE GRUCHY, J. N.: 'Some Applications of the Phase-Sensitive Rectifier Bridge to Electrical Measurements', *Transactions of the Society of Instrument Technology*, 1950, **2**, p. 33.
(27) TAGG, G. F.: 'The Measurement of Earth-Loop Resistance', *Proceedings I.E.E.*, Paper No. 2165 M, September, 1956 (**104** A, p. 215).
(28) ROSCOE, E.: 'Earth Fault Loop Impedance Testing', *Electrical Review*, 27th January, 1956, p. 139.
(29) RUSSELL, R. L., and HODGES, N. W.: 'The Application of Counter-Rotating Fields to Electrical Measuring and Indicating Devices', *Proceedings I.E.E.*, Paper No. 1628, April, 1954 (**101**, Part II, p. 178).
(30) GALL, D. C.: 'Direct-Current Amplifier and Its Application to Industrial Measurements and Control', *Journal I.E.E.*, 1942, **89**, Part II, p. 434.
(31) AXWORTHY, F. R.: 'Electrically Operated Remote Indicators', *Instrument Practice*, 1948, **2**, p. 135.
(32) British Patent No. 589423. 1945.
(33) ARNOLD, A. H. M.: 'Audio-Frequency Power Measurements by Dynamometer Wattmeters', *Proceedings I.E.E.*, Paper No. 1653 M, April, 1954 (**102** B, p. 192).
(34) CHOUDHURY, J. K., and GLYNNE, A.: 'The Use of an Electrostatic Wattmeter for Magnetic-Loss Measurements', *ibid.*, Paper No. 1102, January, 1951 (**98**, Part II, p. 539).
(35) AXWORTHY, F. R., and CHOUDHURY, J. K.: 'The Design and Application of a Portable Electrostatic Wattmeter', *Transactions of the Society of Instrument Technology*, 1952, **4**, p. 57.
(36) British Patent Nos. 617269 and 626700.
(37) MUIRHEAD, LTD.: 'The Importance of Frequency Rate Indication in the Control of the National Grid', *Technique*, October, 1952, No. 4.
(38) BRADSHAW, E., and TANYELOGLU, M. Z.: 'An Instrument for the Measurement of Rate of Change of Power-System Frequencies', *Proceedings I.E.E.*, Paper No. 940, February, 1950 (**97**, Part II, p. 550).
(39) SHOTTER, G. F., and HUTCHINGS, E. E.: 'Potential Indicators for Live Conductor Detection' (E.R.A. Technical Report Ref. F/T133).
(40) SHOTTER, G. F., and HAWKES, H. D.: 'A Precision AC/DC Comparator for Power and Voltage Measurements', *Journal I.E.E.*, 1945, **93**, Part II, p. 314.
(41) ARNOLD, A. H. M.: 'Performance Limits in Electrical Instruments', *Proceedings I.E.E.*, Paper No. 1128, March, 1951 (**98**, Part II, p. 701).

621.372.2

The Institution of Electrical Engineers
Paper No. 2695 M
Sept. 1958

SCIENTIFIC ELECTRICAL MEASURING INSTRUMENTS

A Review of Progress.

By F. C. WIDDIS, B.Sc.(Eng.), Associate Member.

(1) INTRODUCTION

The paper deals with the progress in the design of scientific measuring instruments since the last review was published in 1946.[1] The upsurge in scientific and technological progress during this period has necessitated considerable development in measuring techniques. The rate of progress in any branch of science is determined to a considerable extent by the ability to measure the phenomena under investigation. New instruments have been developed for new applications, and much effort has been devoted to improvements in existing measuring apparatus to cope with the continual demands for improved performance and reliability. In many applications a degree of accuracy is now expected from commercial instruments that has hitherto been achieved only by the most refined laboratory techniques.

It is not possible in the course of a single paper to deal with the whole field of scientific measuring instruments, and the paper is restricted to a consideration of those devices which are primarily of interest to the electrical engineer, or which utilize electrical measuring techniques.

The most important developments in the period that has elapsed since the last review have probably been in connection with nuclear phenomena and their applications, semiconductor devices and electronic instruments.

(2) MATERIALS AND COMPONENTS

(2.1) Resistors and Resistance Materials

The bulk of precision resistance work is still based upon the copper-nickel-manganese alloys, manufactured under several trade names, but commonly known as manganin. The technique for the construction of precision resistors with this alloy is well known and is based upon correct annealing procedure and strain-free construction. The last condition is difficult to satisfy, except in low-value resistors, and while it has long been possible to construct resistors having values up to 10 ohms which would remain constant within a few parts in a million, this reliability cannot be obtained with any certainty with higher-value resistors. A successful strain-free construction suitable for resistors up to 1 000 ohms has been devised by Hall, Gridley and Barber.[3] The bare resistance material is wound into a long close-pitched helix and annealed. The helix is laid loosely in a multi-turn spiral groove cut in a Perspex disc fitted with a cover plate and hermetically sealed. The finished resistor of disc form is comparatively large, but this is not objectionable in a resistor of the highest precision.

Considerable interest has been aroused by the introduction of the nickel-chromium-aluminium alloys with copper or iron.[4] These alloys have resistivities three times that of manganin, a thermal e.m.f. to copper generally less than that of manganin, and a temperature coefficient of resistance which may be controlled by heat treatment at moderate temperatures. A resistance change over a range of temperature of one-tenth of that of manganin can be obtained. Tests indicate that the long-term stability may be comparable with good samples of manganin, though there is some sensitivity to vibration. Owing to the long experience in constructing resistors with manganin, this new material is unlikely to replace it for high-precision d.c. resistors.

The new alloys can be operated over a temperature range up to 100° C, and this, coupled with their high resistivity, is an advantage in a.c. work. The reduced amount of material necessary for a given high resistance permits a substantial reduction in time-constant to be achieved. It is probable that these alloys will be generally adopted for use in high-grade high-resistance boxes. The alloys must be hard-soldered to copper, and this appears to be the only practical difficulty.

(2.2) Instrument Switches

The main changes in design of instrument switches have been due to attempts to simplify and cheapen their production. Moulding of switch parts is now common. One manufacturer is constructing the contacts in decade resistance-box switches from pressings in silver-faced metal, and the same manufacturer has produced a rather novel 100-point switch for precision vernier potentiometers in which the contacts are constructed from copper wire, replacing the conventional gold-silver contact.

The low noise level and low thermal e.m.f. generated in high-quality potentiometer switches have led to their inclusion in motor-driven form into industrial temperature-measuring apparatus.

A new type of instrument switch has been introduced recently in which the switch contact is of sintered silver graphite, one of the so-called Elkonites, and the wiper of silver. This switch has the advantage that it is self-lubricating, and the periodical cleaning and oiling necessary with instrument switches is eliminated. The contact resistance is of the same order as that of a conventional switch in good condition.

(2.3) Materials for Sliding Contacts

Much progress has been made in the development of contact materials; new alloys have been developed for slide-wire contacts such as the copper-silver-gold (625) alloy. This gives excellent results in most types of resistance slide-wire. The development of rhodium-plating techniques now permits the use of this metal for contact surfaces; although very expensive, it appears to be completely stable under every conceivable condition, and ensures absolute reliability of contact even after long periods of idleness. It is particularly useful in conjunction with 625 alloy for light-duty slip rings in electrical measuring circuits associated with rotating machinery. The thermal e.m.f. generated is about 0·095 mV per 100° C. A considerable improvement in noise level and a reduction in the thermal e.m.f. to 0·015 mV per 100° C can be achieved by the use of silver rings with silver-graphite brushes, but this combination is unsuitable for use in sulphur-laden atmospheres.

(3) SEMICONDUCTOR DEVICES

(3.1) Thermistors

The thermally sensitive resistor, or thermistor,[7] is a sintered compound of semiconductors and has a temperature coefficient

Mr. Widdis is at the Northampton College of Advanced Technology, London.

of about 0·04 per deg C. The material is available in beads and discs with a very wide range of resistance values. The resistance/temperature characteristic of the material is reproducible within close limits, and it now forms the basis of many instruments for measuring small temperature changes. The high sensitivity of the material enables temperature changes in the region of 0·001°C to be measured with a relatively simple Wheatstone bridge. Thermistor beads are available having a diameter as small as 0·01 in, and it is apparent that in this form temperature changes in small objects can be readily measured. Other applications include anemometry.

The non-linear relationship between applied voltage and current enables them to be used as voltage-sensitive devices in stabilizers, and as automatic amplitude controls in electronic oscillators and amplifiers.

(3.2) The Hall Effect in Semiconductors

The Hall effect occurs when a transverse magnetic field is applied to a conductor carrying a current. An e.m.f. is set up in the conductor at right angles to the magnetic field and the direction of current flow. This e.m.f. is due to the deflection of the carriers taking part in the conduction process, and is relatively large in semiconductors like germanium. Magnetic-field-strength meters have been developed using this principle, and these may be useful where only limited access to the field exists.[8]

The possibilities of using the Hall effect for power measurement have been studied by Barlow,[9, 10] who has developed semiconductor wattmeters using germanium crystals for use in the audio-frequency range and in waveguides. The output from the germanium crystals is in the microvolt region, which necessitates a rather sensitive detector. The accuracy obtained so far is not very high, although comparable in some cases with existing power-measuring devices. Other semiconductors such as indium antimonide give much larger Hall effects, and their adoption may increase the sensitivity. The method shows considerable promise, and further developments may minimize the disadvantages. This new principle may offer advantages in the measurement of power in heavy-current power systems and in waveguides.

(3.3) Ferroelectrics

The term 'ferroelectrics' is applied to a series of ceramics which exhibit dielectric hysteresis effects closely analogous to magnetic hysteresis in ferromagnetics. Notable among these materials is barium titanate, which develops marked piezo-electric properties after being subjected to a polarizing electric field of the order of 20 kV/cm. This material has a permittivity several hundred times that of quartz, and therefore piezo-electric elements made from it offer a much lower impedance than those previously available. The material is now made use of in a variety of instruments for measuring vibratory strain and acceleration.[11] The sensitivity of the ceramic may be expressed as the field produced per unit pressure in the direction of the polar axis, a typical figure being $1 \cdot 3 \times 10^{-2}$ volt/m per newton/m^2. Very simple accelerometers can be constructed with this material. A typical form consists of a small base plate with an axial stud for screwing to the vibrating body; a polarized disc of barium titanate having a cylindrical mass attached to its other face is bonded to the base plate. Acceleration of the base is transmitted to the mass, which then exerts a pressure on the ceramic. The piezo-electric ceramic generates a charge which is directly proportional to the acceleration at any instant, and has the same waveform as the displacement. A typical design gives 18 mV/g across a high-load impedance. The frequency response is constant between 40 c/s and 10 kc/s, and accelerations up to 1 000g can be withstood.

Very sensitive vibration strain-gauges can be made by silvering opposite faces of wafer-thin strips of this ceramic. The coating on one face is extended round one edge on to a bare region on the other face to facilitate connection. These gauges give an output of about 0·1 volt for a strain of 1 part in 10^6. They are not responsive to static strain, however, and some difficulty arises in calibrating them. As a result, they are more commonly used for the detection of small vibration resonances. The gauges can also be used for exciting vibrations in the body to which they are affixed, by energizing them from an oscillator.

(4) ALTERNATING-CURRENT BRIDGES

A complete new range of a.c. bridges for the measurement of capacitance, inductance and resistance at frequencies up to 100 Mc/s has been developed by the use of inductively coupled ratio arms. The basic principle is not new, having originated with Blumlein in 1928, and much work on radio-frequency bridges using this principle was subsequently carried out at the B.B.C.[12] The full potentialities of the method were indicated by the development of a direct-capacitance aircraft altimeter during the last war.[15] The use of inductively coupled ratio arms permits the measurement of small impedances in the presence of large unwanted stray capacitances. Referring to Fig. 1, T_1 is a voltage transformer whose secondary is tapped

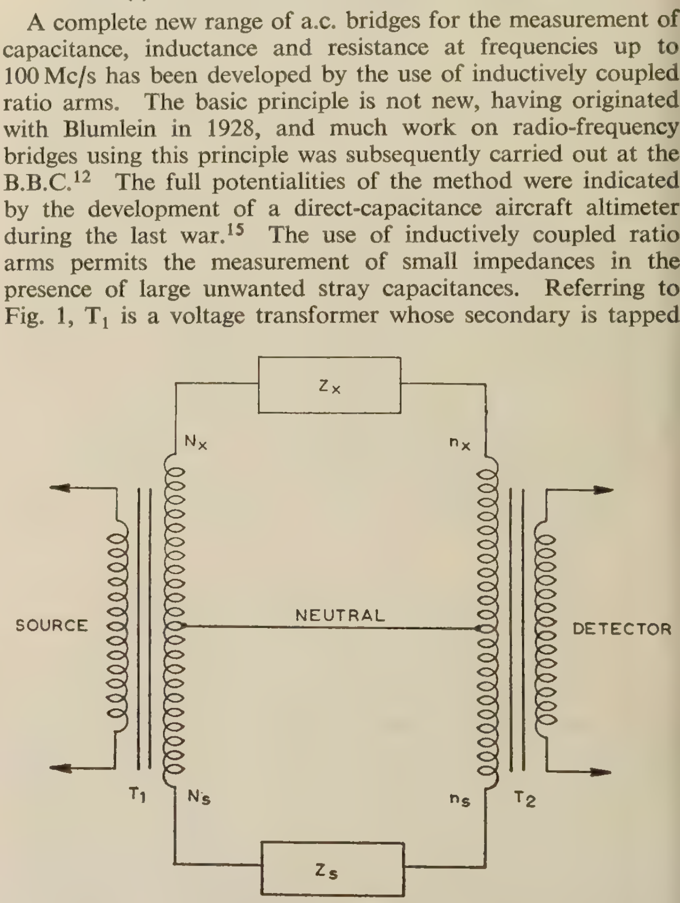


Fig. 1.—A.C. bridge with inductively coupled ratio arms.

to give N_X and N_S turns. T_2 is a current transformer, the primary of which is tapped to give n_X and n_S turns, and the secondary winding is connected to a null detector. Z_S and Z_X are a known and unknown impedance respectively. If Z_S is adjusted to give null indication on the detector, there is then zero flux in the current-transformer core, i.e. the total number of ampere-turns on the core must be zero. It follows that

$$Z_X = \frac{N_X n_X}{N_S n_S} Z_S$$

It is possible by a suitable combination of tappings to obtain a very wide range of measurement in terms of any given standard.

The manner in which stray earth admittances are rendered ineffective may be understood by reference to Fig. 2, which shows an impedance with stray admittances being measured on the bridge. Since there is zero flux in the detector transformer core at balance, there is no potential difference across its wind-

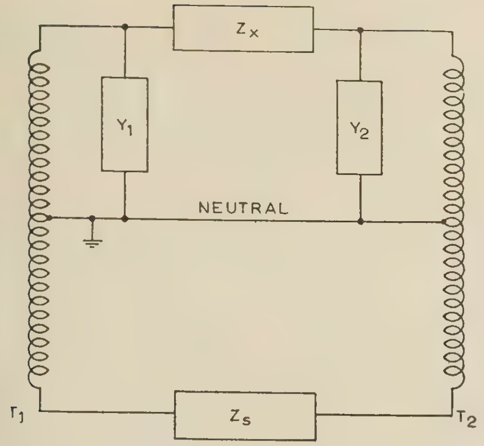


Fig. 2.—Effect of earth admittances on an inductively coupled ratio bridge.

ngs, and hence across y_2, which will therefore not affect the balance. The input admittance y_1 shunts the source transformer and will only affect the bridge balance if the current taken causes a significant drop in the terminal voltage due to leakage reactance and resistance in the transformer windings. This can be counteracted by care in the transformer design. Commercial bridges are available in which capacitances as low as 0·0002 pF can be measured. Useful discussions of these networks will be found in References 12–16.

Many of the difficulties occurring in the use of conventional a.c. bridge circuits at radio frequencies can be eliminated by the use of the twin-T circuit in which source, detector and unknown have a common earth connection. An admittance bridge using this principle has been developed by Woods[17] for use in the range 3–300 Mc/s; it has an error of about 0·2%.

A.C. bridges have also been developed for the assessment of the moisture content of insulating materials.[18] These measure the change of capacitance with frequency over the range 0·05–50 c/s. It is not possible to use a conventional balance detector at very low frequencies, as the time required for balancing is excessive. This difficulty has been resolved by the development of a detector which separates its response into components in phase and in quadrature with the oscillator voltage, and by disposing the bridge circuit so that in-phase and quadrature components can be independently controlled.

A three-terminal twin-T network of resistors and capacitors will reject a single frequency, and this frequency may be varied over a wide range by the use of triple-ganged continuously variable resistors. If such a network is shunted across a single valve stage in an amplifier, the stage gives gain only at the frequency to which the twin-T network is adjusted. This principle has been used in the development of tunable bridge amplifiers for frequency ranges 20 c/s–200 kc/s.

(5) APPARATUS FOR THE PRECISE MEASUREMENT OF ALTERNATING CURRENT, VOLTAGE AND POWER

(5.1) A.C.–D.C. Transfer Devices

Developments in this field have been concentrated upon the provision of standardizing apparatus for calibrating ammeters, voltmeters and wattmeters in the audio-frequency range. An a.c.–d.c. transfer instrument is necessary for such work, and the electrostatic voltmeter has proved to be the most useful and versatile instrument for this purpose in laboratories. The electrostatic voltmeter is believed to have negligible errors at frequencies up to 100 kc/s, and its use in the audio-frequency range is largely determined by the frequency characteristics of the ancillary shunts and multipliers. Much work has been carried out upon the development of these components for use in the audio-frequency range.[23]

An alternative a.c.–d.c. transfer device which has come into prominence as a result of the work of Hermach[24] is the thermal convertor. The thermal convertor has long been used for the measurement of current at high frequencies, but until recently it has not been considered as a transfer device of the highest precision. This is due to certain inherent difficulties, namely the substantial reversal differences which occur when the heater is energized by direct current, the tendency to drift, and lack of understanding of the cause of the a.c.–d.c. transfer errors. It is now known that the errors depend upon the heater material, and units constructed with heaters of the binary alloy 80Ni20Cr may be expected to have transfer errors of less than 1 part in 10^4. The thermal drift is caused by a temperature coefficient which may be as high as 0·002 per deg C. This high temperature coefficient is believed to be due to the presence of unwanted thermo-junctions in the couple connections, and, although these could be eliminated by changes in design, there are serious practical difficulties in such changes. Attempts have been made to eliminate this drift by back-to-back operation of two units, but identity of characteristic is necessary for success, and this is rarely achieved. Drift due to ambient temperature changes is not a difficulty if measurements can be carried out rapidly, and the provision of high-stability electronic power sources has been of great assistance to this end. With such power sources little difficulty is experienced in repeating measurements with a vacuo-thermo-junction to within 5 parts in 10^5.

A simple multi-range volt-ampere convertor, based upon the vacuo-thermo-junction has been developed[25] which will measure alternating current or voltage with an error of $\pm 0{\cdot}05\%$. It has ranges of 15–300 volts and 0·1–5 amp, and an upper frequency limit for 0·05% error between 20 and 100 kc/s, depending upon the range used. This instrument seems to be suitable for most of the accurate alternating current and voltage measurements normally required in a laboratory.

The conventional thermal convertor consists of a straight wire, heated by the passage of the current being measured, with a thermo-junction sensing the mid-point temperature rise. The indirectly heated thermistor has been suggested as an alternative for voltage and current measurement;[26] it has a considerably higher sensitivity but a slower response.

It is possible to measure power in a.c. circuits over a wide frequency range by the use of two thermal convertors, provided that the output e.m.f. of each convertor is proportional to the square of the heater current. A true square-law response cannot be realized in a single convertor, and wattmeters using this principle may be subject to errors of several per cent. The problem has been fully investigated by Hill,[27] who has recently developed a compensated convertor consisting of two units in series, one having a Nichrome heater and the other a platinum heater. With this combination a true square-law response may be obtained. A new wattmeter has been developed, using these compensated convertors, with an overall instrument error for all conditions of load and power factor of 0·1% over the frequency range 200 c/s–10 kc/s, and 0·3% for the range 50 c/s–30 kc/s.

(5.2) A.C. Potentiometers

There are many types of a.c. measurement where it is desirable that the measuring apparatus should absorb negligible power from the circuit. The a.c. potentiometer is extensively used for such measurements, and it will operate satisfactorily in the audio-frequency range if a vacuo-thermo-junction is used as the

transfer device. At these frequencies unwanted capacitance currents in the potentiometer can cause serious errors, but these errors can be largely eliminated by careful shielding.

(5.3) Indicating Instruments incorporating Feedback Amplifiers

Electronic negative-feedback amplifiers are now being used in conjunction with high-grade instruments as impedance-changing and amplifying devices. Provided there is sufficient feedback, the gain of such an amplifier is unaffected by changes in valves and other components, and may be determined solely by a high-grade wire-wound resistor. Such amplifiers are combined with moving-iron or electrodynamic milliammeters to form accurate measuring instruments having very high input impedances. Electrodynamic wattmeters have been fitted with feedback amplifiers for use in the audio-frequency range with minimum circuit loading. An advantageous feature is that the amplifier forces the voltage-coil current to be in phase and proportional to the amplifier input, thus eliminating effects due to the inductance of the voltage coil.[28]

(5.4) Stabilized Power Supplies

The extension of precise voltage, current and power measurement into the audio-frequency range has necessitated the provision of high-stability variable-frequency power supplies. This requirement has been met by the development of stabilized electronic-oscillator amplifier sets. The Wien-bridge *RC* oscillator is particularly suitable for such applications because of the ease with which it can be stabilized by a non-linear element such as a lamp or a thermistor. Such oscillators can be designed to maintain their output level constant within $\pm 0{\cdot}01\%$ for short periods; in addition the waveform is good, as a total harmonic content in the region of $0{\cdot}1\%$ can be attained.

Output powers of the order of hundreds of watts have been produced by the attachment of amplifiers to the oscillator output. Phase-shifting networks in conjunction with a number of amplifiers have been used to produce two outputs continuously variable in phase suitable for wattmeter testing. Three-phase outputs are also available. The output stability of a typical oscillator-amplifier set is about $0{\cdot}01\%$ over one minute, with a harmonic content of $0{\cdot}3\%$. It is necessary to provide the amplifier and oscillator with stabilized anode and heater supplies unless a large amount of negative feedback is incorporated.[29]

An interesting development has been the production of an oscillator having a purely sinusoidal voltage waveform.[30] This oscillator has been used in the determination of the a.c.–d.c. transfer error of an electrostatic voltmeter by making use of the fact that the effective value of a pure sine wave may be determined accurately from measurements of either its mean or peak value.

Electronic stabilizers are now available for stabilizing the 50 c/s mains supply. A noteworthy example is that due to Patchett,[31] which in a typical case will give an undistorted output up to 500 watts, stable to within $0{\cdot}01\%$. The performance is independent of the load power factor. This particular stabilizer uses a thermistor bridge as the voltage-sensitive element.

Electronic stabilizers are also being produced using saturated diodes as the voltage-sensitive element.

(6) NUCLEAR DEVICES

(6.1) Radiation-Measuring Instruments

Measurements in nuclear apparatus and plant have necessitated the development of new types of instrument. These new instruments are either for radiation measurement or utilize radioactive radiations.

Radiations from radioactive substances may consist of nuclear particles (α-particles, β-particles, protons, neutrons and mesons) or very short-wave radiations (γ-rays). Charged particles produce ionization or excitation of material through which they pass and are normally detected by gas-ionization devices such as ionization chambers, proportional counters and Geiger-Müller counters. The scintillation counter is an alternative form of detector in which the charged particle impinges on a phosphor to give a flash of light.

Most of these devices originated many years ago, and recent development has been concentrated upon improvements in performance, such as the introduction of the self-quenching Geiger–Müller counter. The combination of a phosphor with a photo-multiplier tube has made the scintillation counter a most versatile radiation-measuring device, giving an output pulse proportional to the energy of the incident particle. A particular advantage of the scintillation counter is its very short resolving time, which is usually less than 1 microsec, but depends upon the phosphor used.

The problem of detecting uncharged particles (neutrons) is more difficult, and they can only be detected by causing them to interact with matter emitting secondary charged particles. Slow-neutron flux is normally measured with an ionization chamber filled with boron trifluoride gas, or having electrodes coated with solid boron and the chamber filled with hydrogen. The neutrons react with boron 10 nuclei producing α-particles and charged lithium nuclei which cause ionization of the gas. There are two common forms of pulse-counting neutron detectors; the first is the boron-trifluoride-filled proportional counter, and the second the fission counter in which neutrons cause fission in a thin coating of uranium^{235}. The fission fragments cause ionization of the filling gas. Fast neutrons may be detected with a fission chamber or from the ionization produced by recoil protons due to elastic collisions in hydrogenous material.[35, 36]

The output current from an ionization chamber is usually measured with an electrometer d.c. amplifier, the complete instrument being termed a 'ratemeter'. The amplifier is usually designed to have a logarithmic response. The output pulses from counters are counted with a scaler similar to that mentioned in Section 7.6. When a particular radiation is being studied, problems arise due to the presence of other radiations which give a background count. Discrimination is effected by making use of the differing penetrating powers of the various forms of radiation, and by the introduction of a pulse-amplitude discriminator which is simply a d.c. bias permitting pulses above a certain level to be counted. The coincidence circuit is also useful in discrimination, permitting a pulse to pass only when actuated at the same time by different counters.[37] Where the electronic circuits associated with radiation measurement deal with pulses, cold-cathode tubes are very suitable, particularly in portable apparatus. They have the advantages of power economy, robustness and reliability, and sometimes economy in components.

The practical application of radiation techniques which is of most interest to the engineer is the thickness gauge. The β-ray thickness gauge is now coming into general use for the measurement of thickness (or weight per unit area) of sheet materials such as paper, board, plastics, metals, etc. β-radiation from a radioactive isotope is directed towards a detector which measures the strength of radiation. When a sheet of material is placed between the isotope and the detector, some of the radiation is absorbed, and the decrease in radiation is proportional to the weight per unit area of the material. The thickness of coatings can be determined by measuring back-scatter.

There is a limit to the depth of penetration of β rays, and the γ-ray thickness gauge is a new device which functions with moderately thick materials. Such gauges have been developed

for monitoring the thickness of hot steel strip in the range 0·05–0·30 in.[39]

(6.2) Nuclear Magnetic Resonance

A precise method for the measurement of magnetic fields has been devised, based upon the phenomenon of nuclear magnetic resonance.[41]

Nearly all atomic nuclei possess spin and magnetic moments, and in the presence of an external magnetic field the direction of spin processes around the direction of the field. This phenomenon is known as the Larmor precession and occurs at the Larmor precession frequency. Each nucleus has a specific constant relationship between the Larmor frequency and the flux density of the applied magnetic field. Determination of the Larmor frequency enables the magnetic field strength to be measured with precision. The relationship is dependent upon the condition in which the nuclei are present, i.e. in elementary form or combined in molecules. For the proton in distilled water $B = 2{\cdot}348 \times 10^{-8}f$, where f is the Larmor precession frequency in cycles per second and B is the magnetic field density in webers per square metre. It can be seen that when $B = 1\,\mathrm{Wb/m^2}$ resonance occurs at 42·85 Mc/s, which is a convenient radio frequency.

In a practical instrument[42] a proton sample of distilled water in a glass cell 1–2 cm long by 0·5 cm inside diameter is mounted with a small coil and inserted in the magnetic field to be measured with the axis of the coil at right angles to the direction of the field. Radio-frequency oscillations are set up in the coil, and when their frequency coincides with the Larmor precessional frequency of the protons, transitions occur which result in an absorption of energy from the r.f. field. A weak magnetic field varying at 50 c/s is superimposed on the field being measured by inducing coils, and, by a suitable detector, the absorption in energy from the oscillator when the correct frequency is reached is shown as a V-shaped dip in a horizontal c.r.o. trace.

An alternative form of detection utilizes a second coil placed in a position of zero mutual inductance with respect to the exciting coil and connected to a receiver. When the frequency of the oscillator reaches the Larmor frequency there is radiation from the sample, and it is detected by the receiver. It is necessary to keep the receiver tuned to the oscillator.

An error of $\pm 0{\cdot}2\%$ is claimed for one instrument, but it is possible to reduce the errors to the order of one part in 10^6. It is essential that the fields being measured are constant; non-homogeneity causes broadening of the resonance line, and the resonance becomes unrecognizable when the non-homogeneity over the sample exceeds 1 part in 500.

(6.3) The Caesium Resonator

The magnetic properties of an atom depend upon the angular momentum of the extra-nuclear electrons, their spin, and the magnetic moment of the nucleus. If atoms are passed through an alternating field at a particular frequency, there may be a change of energy state associated with a change in polarity of the magnetic dipoles. The full theory of this phenomenon is outside the scope of this review but is dealt with fully in the literature. The principle has been used by Essen and Parry[43] to develop a new standard of frequency with an extemely high accuracy.

The new standard consists of an evacuated tube with an oven at one end containing a caesium source. A beam of caesium atoms emerges from a slit in the oven and passes through a constant deflecting magnetic field; the beam then passes through an exciting r.f. field superimposed upon a weak constant magnetic field. If the frequency of the r.f. field is at the critical value, the magnetic polarity of some of the atoms is changed. This change of polarity is detected by passing the beam through a second constant magnetic field which deflects those atoms which have undergone transition on to a detector. The detector is a tungsten wire heated to 1 000° C, and the caesium atoms striking it are emitted as positive ions which are collected by a curved plate maintained at −20 volts relative to the wire. The current flowing is measured with the aid of a 10^{10}-ohm resistor and a vibrating-reed electrometer amplifier.

The resonance occurs at approximately 9 192 Mc/s and has a frequency bandwidth at half-deflection of 350 c/s. The peak can be set to about ± 1 c/s, thus the frequency may be defined to within ± 1 part in 10^{10}.

This resonator is used to calibrate quartz clocks and has the advantage that, once the exact frequency is determined in terms of the second of ephemeris time, it provides a standard independent of astronomical measurements. Compared with astronomical methods it is believed to give an improvement in accuracy of over 100 times when calibrating quartz clocks.

(7) ELECTRONIC INSTRUMENTS

Improvement in electronic instrumentation has been a notable development since the last review. The electronic instrument was previously a relatively crude and unreliable device, but it is now an indispensable laboratory tool. This metamorphosis has been brought about by the development of reliable components, and by a fuller understanding of the principles of negative-feedback amplifier design.

The use of the transistor may lead to revolutionary changes in the design of electronic instruments, but it is too novel for a critical assessment of its effect upon electronic instrument design. It is safe to assume that it will eventually replace the thermionic valve in many instruments.[53] The possible advantages arise from the very small physical dimensions and low power consumption, which may result in a substantial reduction in instrument size. Existing electronic instruments are usually bulky devices, dissipating appreciable power.

Printed-circuit techniques are now common in mass-produced electronic apparatus, and these techniques are being adopted in the relatively small-batch production of some electronic instruments.

(7.1) Direct-Current Amplifiers

(7.1.1) D.C. Amplifiers for Low-Resistance Input Circuits.

A wide variety of scientific measurements require the amplification of small d.c. signals which may originate from low-resistance sources such as thermocouples or thermopiles. The design of suitable electronic d.c. amplifiers for these purposes is a difficult problem owing to the difficulties in eliminating drift. For many purposes adequate performance can be achieved by the construction of balanced amplifiers with the valves arranged in long-tailed pairs, but this is not feasible if high levels of gain are required. The output signals from low-impedance sources such as bolometers are only of the order of microvolts. Present practice converts the d.c. input signal to a low-frequency a.c. signal by means of motor-driven contactors. The a.c. signal is then amplified and re-converted to a d.c. signal by a synchronized contactor at the output. Commercial amplifiers using this principle are available which will detect $0{\cdot}01\,\mu$V, and which will not drift more than $0{\cdot}01\,\mu$V over 24 hours. The contactors have gold-alloy contacts, are surrounded by high-permeability alloy screens, and are actuated at frequencies between 8 and 80 c/s by motor-driven shafts. When dealing with larger signals in the region of several hundred microvolts, the contact modulation technique is also used, but at these levels the contactor may consist of a vibrating reed.[44, 47]

These contact-modulated amplifiers are restricted to d.c. signals and very low-frequency a.c. signals. In many applications the requirement is for an amplifier having a wide frequency response extending down to zero frequency. This requirement is met by the drift-corrected direct-coupled amplifier. In this device, a point is chosen on an output network in a direct-coupled amplifier; it will have the same potential as the short-circuited input terminals in the absence of drift in the amplifier. The potential difference between these points and the input terminal is continuously monitored by an auxiliary-contact modulated amplifier which injects a correcting signal into the main amplifier. The zero drift can be reduced to a few microvolts per hour, as opposed to about 1 mV for the normal balanced amplifier.

An alternative approach to the problem of wide frequency range coupled with d.c. response has been suggested by Buckerfield,[45] who has developed a contact-modulated amplifier connected in parallel with a wide-band a.c. coupled amplifier.

(7.1.2) **D.C. Amplifiers for High-Resistance Input Circuits.**

The measurement of small direct voltages in high-resistance circuits necessitates the use of a d.c. amplifier with an extremely high input resistance. Electrometer valves have been in use in the input stages of such amplifiers for many years. These electrometer valves have long been available having grid currents as low as 10^{-15} amp, and they are now constructed in double form using a common cathode in order to minimize drift. A typical balanced arrangement might drift about 100μV in several hours. A more recent development is the subminiature electrometer triode. This has a grid current of about 10^{-13} amp, which is sufficiently low for a wide variety of instrument applications, notably in pH work. Its small size and low power consumption have led to notable design improvements in pH-meters.

The electrometer pentode—an indirectly heated valve with a grid current of less than 10^{-11} amp—is now extensively used in commercial direct-indicating pH-meters.

The measurement of very small currents from ionization chambers is of importance in nuclear physics. The procedure is to pass the current through a high resistance and then measure the potential difference across the resistor with the aid of an electrometer amplifier. Resistors are now available with resistances up to 10^{12} ohms, having a stability at room temperature of better than 2% per month. These resistors are normally vacuum-sealed in glass, and surface leakage is minimized by coating the glass envelope with a water-repellent silicone resin.

The vibrating-reed electrometer is an alternative to the electrometer valve for measurements in high-resistance circuits.[50, 51, 52] The direct voltage to be measured is applied through a high resistance to a small capacitor, one plate of which is subjected to continuous mechanical vibration by a solenoid energized with alternating current. Owing to the alterations in capacitance, an alternating voltage appears across the plates of the capacitor, and this voltage may be amplified by an a.c. amplifier. The input resistance can be very high, but some difficulty has been encountered in the past in maintaining zero stability owing to variations in the contact potential between the plates. This difficulty has now been overcome, and these vibrating-reed electrometers are now available commercially with input resistances up to 10^{16} ohms, and with a zero stability of about 10μV in 2 hours. The stability is claimed to be about 10 times that of the best electrometer-valve arrangement.

(7.2) High-Resistance Measurement

High resistances are measured conventionally by applying a high voltage from a battery to the specimen, and then measuring the current with a sensitive galvanometer. This form of measurement can now be effected by an electronic instrument which supplies the necessary test voltage and indicates the value of resistance on a meter. The range of the instrument is from 3×10^5 to 2×10^{13} ohms. The test voltage, at either 85 or 500 volts, is obtained from an electronically stabilized mains-operated power unit, and the leakage current is measured by passing it through a stable high-value resistor. The potential difference developed across the resistance is amplified by an electrometer amplifier provided with overall negative feedback. The output from the amplifier actuates a meter calibrated in resistance values.

It is probable that this type of instrument will be the basis of future laboratory insulation measurements. There are three major difficulties in such a design which appear to have been satisfactorily surmounted: a high degree of stability must be achieved in the high-voltage supply, particularly when measuring resistance in the presence of large capacitance; the d.c. amplifier must have a high input resistance, be drift free, and stable in gain; and the high-value measuring resistor must be stable.

(7.3) Cathode-Ray Oscillographs

The most important development in this field is probably the post-deflection acceleration tube. The electron beam passes through the main accelerating electrostatic field after the X and Y deflecting plates; as a result a very high deflection sensitivity is achieved, coupled with a high writing speed. In one design the accelerating field is obtained by means of a close-pitched helical track on the body of the tube. Tube sensitivities as high as 1 cm/volt are attainable. The high sensitivity of these tubes simplifies the design of the associated amplifiers and has resulted in a marked improvement in oscilloscope performance. One commercial model now has a bandwidth from zero frequency to 30 Mc/s, with a sensitivity of 0·05 volt/cm, and a rise time of 12 millimicrosec, a performance invaluable for the study of transient phenomena at low signal levels.

Continuous improvements have been effected in tube construction in order to improve linearity, and measuring oscilloscopes now have flat ground and polished faces.

The measurement of the high-voltage transients occurring in the surge testing of electric power equipment is now generally carried out with sealed cathode-ray tubes operating at about 10 kV, as distinct from the earlier continuously evacuated tube. The signal level in such measurements is high, and the maximum recording frequency is determined by the tube resolution. In existing commercial models the upper frequency limit is about 100 Mc/s, the ultimate resolution obtained on a photographic record being about 0·6 millimicrosec.

The method of delineating a recurrent waveform by the measurement of its instantaneous value at different points during recurrences is well known. The recent application of electronic techniques to this principle has resulted in the development of a cathode-ray oscillograph for monitoring high-speed recurrent waveforms. The instrument will display without distortion recurrent waveforms having frequency components up to 300 Mc/s and amplitudes as small as 0·1 volt. The shortest time-scale has a length of 0·05 microsec.

(7.4) Valve Voltmeters

The conventional form of valve voltmeter consists of a diode probe associated with a d.c. amplifier. The upper frequency limit is determined by the probe design and now approaches 1 000 Mc/s. It is customary to use miniature vacuum diodes in the probe, but crystal diodes are also employed. In the past, one difficulty with valve voltmeters has been due to valve drift; this can be substantially reduced by the use of double valves with

common heaters in a differential connection. This technique is now generally applied in valve voltmeters, both in the vacuum diode probe and the following d.c. amplifiers, and recent designs are almost entirely drift-free.

Sensitive valve voltmeters for use at lower frequencies have been developed in which the signal is amplified by a stable negative-feedback amplifier and indicated on a meter provided with crystal rectifiers.

The in-phase component of an unknown voltage can be measured by applying it to one coil of a wattmeter movement and supplying the other coil from a reference voltage. The quadrature component can then be measured by rotating the reference voltage through 90° by a phase-shifting network. This technique is well known, but a new phase-sensitive voltmeter has been introduced in which the wattmeter movement is supplanted by two thermocouple wattmeters, each actuated by an electronic amplifier. The two components are simultaneously displayed, and a typical instrument covers a frequency range of 20 c/s–20 kc/s, with voltage ranges of 15 mV–15 volts. The input impedance is high, and the device is suitable for measurement of the transmission characteristics of networks and amplifiers.

(7.5) Impedance and Phase-Measuring

The phase angle between two equal voltages is given by $V_T = 2V_1 \cos \phi/2$, where V_T and V_1 are the sum and component voltages respectively. If two voltages differing in phase are adjusted to equality in magnitude by means of variable-gain amplifiers, and the sum indicated on a rectifier-type instrument, the instrument can be calibrated in phase angle. This principle has been used in some recently developed phase-meters which can measure the phase angle between two voltages differing considerably in magnitude over a frequency range of 2 c/s–100 kc/s.

The principle of the phase meter has also been used in the impedance-angle meter, which measures impedance and phase angle of electrical networks over a range of 3–500 000 ohms and $\pm 90°$. An alternating voltage is applied to the unknown impedance in series with a variable calibrated resistor, and the potential difference across each is measured by a valve voltmeter. The resistor is adjusted until both are equal, and the unknown impedance is then equal in magnitude to the value of the resistor. The phase angle is measured in a manner similar to that used in the phase meter. The frequency range is 30 c/s–20 kc/s, with an accuracy within $\pm 2\%$ and $\pm 2°$.

(7.6) Scalers, Timing and Frequency-Measuring Devices

The so-called particle counters used in the detection of radioactive radiations are actually detectors which give a series of voltage pulses related to the strength of the incident radiation. The output from these devices is sometimes fed to a counting device termed a scaler, which indicates the total number of pulses reaching it. The high counting speed necessary is achieved with electronic circuits, and the display is normally in digital form, the numbers being illuminated by miniature neon tubes. The resolution time of such scalers is about 1 microsec, and a 6-unit display will give a count up to 10^6.

The scaler, in a modified form, can readily be used for precise frequency measurement by incorporating gating circuits actuated from a crystal-controlled oscillator. The gating circuits switch on the scaler for a precise time-interval, and the number of cycles of the test frequency occurring in that interval is counted. Sampling times between 0·01 and 10 sec give an overall frequency range of 10 c/s–1 Mc/s with an error of 1 count. Small time-intervals can be measured with a high precision by running the scaler from a 1 Mc/s crystal oscillator and starting and stopping the count by externally derived pulses.

The Dekatron,[56] a scale-of-ten gasfilled cold-cathode counting tube introduced some years ago, is useful for counting frequencies up to about 4 kc/s. It is understood that later forms of this device operate at frequencies up to 20 kc/s. A single Dekatron will give a succession of output pulses at one-tenth of the input frequency, thus acting as a simple frequency subdivider. This property has been used by one instrument manufacturer to derive a range of time-marking intervals for oscillographic records from an electrically maintained fork.

Many forms of time-interval measurement, notably those associated with relay timing, fuses and switch operation, do not require a high degree of accuracy, but the interval to be measured may be a few milliseconds. This class of measurement is adequately effected by electronic time-interval meters based upon the measurement of the charge accumulated in a capacitor during the interval. Typical ranges of existing commercial instruments are 0–4 millisec to 10 sec, and 0–40 microsec to 10 millisec with an error between $\pm 1\%$ and $\pm 5\%$.[58]

(7.7) Wave Analysers

The heterodyne wave analyser covering the range 16 c/s–16 kc/s has, until recently, been the only satisfactory instrument in this field. Since it operates with a constant bandwidth of about 4 c/s, its performance falls off at the low-frequency end of the range.

The study of mechanical vibrations has become of considerable importance in recent years. An analyser for this work must deal with very low frequencies of the order of a few cycles per second. In addition, the frequencies may fluctuate appreciably, and the waveform may contain several unrelated components differing only slightly in frequency. The Pametrada (Parsons and Marine Engineering Turbine Research and Development Association) wave analyser has been developed to meet this requirement, and it operates on a different principle from the heterodyne analyser. It uses selective amplifiers which are tuned by *RC* networks to the frequency of the selected wave component, its magnitude then being indicated on an output meter. It is, in effect, a tunable filter with a range of 19 c/s–20 kc/s, which can be extended down to 2 c/s. The particular advantage of this arrangement is that it offers a constant percentage bandwidth at all frequencies, as compared with the constant bandwidth of the heterodyne analyser. This difference is rather important when dealing with low frequencies.

Simplified forms of the Pametrada analyser have been developed which are particularly useful as bridge detectors in cases where there is appreciable harmonic content in the bridge output.

(8) ELECTRICAL APPARATUS FOR MEASURING NON-ELECTRICAL QUANTITIES

(8.1) Electrochemical Measuring Instruments

The pH-meter is the most common electrochemical measuring instrument. The design of this instrument has undergone some improvements, such as the introduction of the sub-miniature electrometer valve. This has resulted in an increase in the input resistance and a reduction in the power consumption. Direct-reading pH-meters with automatic temperature compensation are now fairly common and are made with an error as low as $\pm 0{\cdot}02$ pH. The usual technique bases the design upon a balanced single or multi-stage d.c. amplifier, using an indirectly heated electrometer pentode at the input.

A development arising from these instruments is the automatic titrimeter for the automatic performance of routine titrations in the laboratory. The required pH-value is set by a dial on the instrument, and the difference between the setting and the

signal from a glass electrode system in the solution actuates a solenoid-operated tap in the burette, which is closed when the end-point is reached.

Automatic coulometric titrimetry is a technique which has aroused interest.[61] In this process an electric current is passed through a solution to liberate ions, which are then used as the reagent. In a typical application a solution of Na_2SO_4 is electrolysed and converted into H_2SO_4 and NaOH. The H_2SO_4 is used as the titrant. The actual quantity of acid used can be measured in terms of coulombs, since it is directly proportional to the current passed through the cell and to the total time of electrolysis. In practice, the reagent is fed into a coulometric cell, where electrolysis takes place. There is an outlet tube at each end of the cell, sealed with a sintered glass disc in which is incorporated a fine wire to form the anode or cathode. One outlet tube delivers H_2SO_4 into the test solution. The number of coulombs passed is integrated by a motor fitted with a cyclic counter. The operation can be entirely automatic, and the technique is claimed to be particularly useful in micro-titration.

The technique of chemical analysis known as polarography depends upon the fact that the application of a slowly changing direct potential between two electrodes in a solution will cause each element in the solution to plate out at the characteristic potential termed the 'half-wave potential'. The electrode current rises in a series of steps, and the height of each step is related to the concentration of its associated element. Contamination of the electrode surfaces is avoided by the use of dropping-mercury electrodes. Existing polarographs apply a rising direct potential to the electrodes, and the electrode current is recorded. A variety of recording methods are in use, some instruments giving ink records and others photographic records. The cathode-ray oscilloscope is also used. Recent developments have been concentrated upon improving the definition of these instruments by recording the derivative of the current/voltage curve. A discussion of the relative merits of the latest techniques will be found in Reference 63.

(8.2) Electro-Medical Apparatus

Certain electro-medical techniques, such as electro-cardiography and encephalography, have been well established for many years. The introduction of electronic amplifiers into this field has led to the development of direct ink-writing instruments as an alternative to photographic types. The recorders for some of these applications must be able to respond to frequencies in the region of 50 c/s. One type of direct-writing pen recorder suitable for these frequencies uses a dry electrolytic paper, which is marked by the passage of electric current from the pen. Another has a heated-wire loop moving in proximity to a wax-coated paper, the heat melting the wax and exposing the black backing.

The ability of the cathode-ray oscilloscope to give two-dimensional displays has led to the technique of vector cardiography, in which the signals from electrodes are simultaneously applied to the X and Y plates, giving a vector loop. The method is claimed to have diagnostic advantages.

The dye-dilution curve recorder is a new instrument. In this device cardiac output is determined by injecting a dye into the blood stream and recording the dye concentration as it passes through the capillaries of the ear. The sensitive element consists of two photocells and a light source which are attached to the lobe of the ear. The photocells are mounted side by side, one being sensitive to infra-red and the other to red rays.

Medical applications of the pH-meter have been extended by the introduction of special glass electrodes for determining the pH-value of soft tissues such as brain tissue. Another type of glass electrode, together with the reference electrode, can be swallowed in order to measure the pH-value of the stomach content.

(8.3) Temperature-Measuring Apparatus

A new design for a platinum resistance thermometer with a stability of $\pm 0{\cdot}001^\circ$C has been introduced by Barber.[64] A strain-free construction has been adopted, the resistance element consisting of a helix of 0·05 mm platinum wire mounted in a fine tube which is then bent into a hairpin form and enclosed in an outer sheath of 6 mm diameter. The whole thermometer is filled with dry air and hermetically sealed. The lag constant is 5 sec and the normal temperature range is from -180 to $+600^\circ$C.

The high performance of this thermometer, together with demands for increased accuracy in the maintenance of the International Temperature Scale, has necessitated the development of an improved Smith bridge type 3. For platinum resistance thermometry the Smith bridge type 3 appears to provide the most satisfactory measurement process. This arises from the inherently high operating speed, and because it is the only bridge which will permit measurements to be made to an accuracy within $0{\cdot}0001^\circ$C with a single setting of the dials. The new bridge utilizes the special strain-free resistances referred to in Section 2.1 and is temperature-controlled at 27° C by circulating air.[65] It has been suggested that the adoption of a modified circuit due to Gautier[66] would lead to further improvement.

In general, only minor detail improvements have been made in optical and radiation pyrometers. Special lamps with optically flat faces on the glass envelope are now available for use in optical pyrometers, resulting in considerable improvement in calibration.

Temperatures below 1 000° C can normally be measured with the aid of thermocouples, but in some cases a radiation pyrometry technique is preferable. The use of a lead-sulphide photo-conductive cell permits radiation measurements to be carried out at temperatures as low as 120° C. These cells have a peak response in the region of 2·5 microns; they are rather unstable, however, and it is necessary to chop the incident radiation and use a tuned amplifier to obtain reasonable reproducibility and sensitivity. In an instrument developed by Barber and Pyatt[68] the radiation from the test surface is compared with that from a lamp at a comparatively high temperature behind an absorption screen, the lamp current being adjusted until the two are equal.

A particularly interesting application of the lead-sulphide cell is due to Parker and Marshall,[67] who have successfully used it for the measurement of the surface temperature of rotating brake-drums during braking. In particular, the small response-time of these cells, of the order of a few microseconds, assists this work, since it enables rapid variations of surface temperatures to be displayed on the screen of a cathode-ray oscilloscope. The technique can clearly be extended to the measurement of the temperature of any moving surface.

The thermistor has been previously referred to, in Section 3.1. The small sizes available, and the high sensitivity of this material, has led to its use in many temperature-measuring applications. Temperature changes as little as $0{\cdot}001^\circ$C may be readily detected with simple auxiliary apparatus. Tiny beads of this material have recently been used for measuring the temperature inside the buds of growing plants.

(8.4) Civil and Mechanical Engineering Applications

Some applications in this field have already been dealt with in previous Sections; in particular the measurement of thickness by radiation has been described. Radiation sources are also used extensively for the equivalent to X-ray examination of welds in

pipes, and in the examination of prestressed-concrete structures for cavities.

The electrical-resistance-wire strain gauge is now in general use and has proved to be invaluable in all structural work. All the earlier gauges consisted of fine wire grids affixed to paper backing, but in recent years a new type of gauge has been introduced consisting of a metal-foil grid on an epoxy-ethylene backing. A printed-circuit technique is used in the construction of this gauge. The new gauge is somewhat larger than the wire gauges, but a wide variety of gauge forms is possible, and the maximum operating current is higher. Many weighing and pressure-measuring devices have been developed based upon the resistance strain gauge. Much attention is being given to the problem of measuring strain at elevated temperatures.

The technique of flaw location by ultrasonics is now well established, and this technique has been extended to the measurement of the density of concrete by measuring the velocity of propagation of ultrasonic waves through known thicknesses of the material. The density of concrete is related to its mechanical strength.

Small mechanical displacements are now generally measured by either inductance or capacitance transducers. An extremely high sensitivity can be obtained with such devices,[75] and a useful review of this field has been given by Woodcock.[77]

(9) FUTURE TRENDS

There will undoubtedly be a further expansion in the field of electronic instrumentation, and efforts will be directed towards improvement in the general reliability of electronic devices.

The multiplicity of instruments now used in experimental work introduces a serious space problem, and it is desirable that efforts should be made to effect a substantial reduction in instrument size without sacrifice of performance. The transistor may enable this to be done in the case of electronic instruments.

Automatic measuring and indicating instruments are common in industrial process work, and in some cases an error of the order of 0·1% is attainable. Some recent industrial instruments display the measured quantity in digital form, which can give a very high reading accuracy. It is probable that this technique may be generally extended to laboratory instruments; self-balancing a.c. bridges have already been constructed for routine network analysis. This may be of importance where the readings are made by semi-illiterate operatives in under-developed countries.

(10) ACKNOWLEDGMENTS

The author wishes to acknowledge the assistance he has received from many scientific-instrument manufacturers and research organizations in the compilation of this report. He is solely responsible for all opinions expressed therein.

(11) BIBLIOGRAPHY

(1) Gall, D. C.: 'Electrical Measuring Instruments. Part 2—Scientific Instruments', *Journal I.E.E.*, **93**, Part I, 1946, p. 352.
(2) Schulze, A.: 'Die neure Entwicklung von Normalwiderständen', *Elektrotechnische Zeitschrift*, 1954, **75**, p. 547.
(3) Barber, C. R., Gridley, A., and Hall, J. A.: 'A Design for Standard Resistance Coils', *Journal of Scientific Instruments*, 1952, **29**, p. 65.
(4) Arnold, A. H. M.: 'Nickel-Chromium-Aluminium-Copper Resistance Wire', *Proceedings I.E.E.*, Paper No. 2084 M, July, 1956 (**103** B, p. 439).
(5) Clothier, W. K.: 'A Fixed Gas-Dielectric Capacitor of High Stability', *ibid.*, Paper No. 1679 M, 1954 (**101**, Part II, p. 453).
(6) Griffiths, W. H. F.: 'Decade Air Condenser', *The Engineer*, 1956, **292**, p. 691 and p. 728.
(7) Becker, J. A., Green, C. B., and Pearson, G. L.: 'Properties and Uses of Thermistors—Thermally Sensitive Resistors', *Bell System Technical Journal*, 1947, **26**, p. 170.
(8) Saker, E. W., Cunnell, F. A., and Edmond, J. T.: 'Indium Antimonide as a Fluxmeter Material', *British Journal of Applied Physics*, 1955, **6**, p. 217.
(9) Barlow, H. E. M.: 'The Application of the Hall Effect in a Semi-Conductor to the Measurement of Power in an Electromagnetic Field', *Proceedings I.E.E.*, Paper No. 1654 M, June, 1954 (**102** B, p. 179).
(10) Barlow, H. E. M.: 'The Design of Semi-Conductor Wattmeters for Power-Frequency and Audio-Frequency Applications', *ibid.*, Paper No. 1778 M, November, 1954 (**102** B, p. 186).
(11) Kell, R. C., Luck, G. A., and Thomas, L. A.: 'The Use of Ferro-electric Ceramics for Vibration Analysis', *Journal of Scientific Instruments*, July, 1957, **34**, p. 271.
(12) Kirke, H. L.: 'Radio-Frequency Bridges', Radio Section Chairman's Address, *Journal I.E.E.*, 1945, **92**, Part III, p. 2.
(13) Calvert, R.: 'A New Technique in Bridge Measurements', *Electronic Engineering*, 1948, **20**, p. 28.
(14) Clark, H. A. M., and Vanderlyn, P. B.: 'Double-Ratio A.C. Bridges with Inductively-Coupled Ratio Arms', *Proceedings I.E.E.*, Paper No. 742 M, 1949 (**96**, Part III, p. 189).
(15) Watton, W. L., and Pemberton, M. E.: 'A Direct-Capacitance Aircraft Altimeter', *ibid.*, Paper No. 743 M, January, 1949 (**96**, Part III, p. 203).
(16) Oatley, C. W., and Yates, J. G.: 'Bridges with Coupled Inductive Ratio Arms as Precision Instruments for the Comparison of Laboratory Standards of Resistance or Capacitance', *ibid.*, Paper No. 1631 M, March, 1954 (**101**, Part III, p. 91).
(17) Woods, D.: 'A Precision Dual Admittance Bridge for the Range 3 Mc/s to 300 Mc/s', National Physical Laboratory Symposium on Precision Electrical Measurements 1954, H.M. Stationery Office, Paper No. 4.
(18) Mole, G., and Smith, D. C. G.: 'A Bridge for the Measurement of the Admittance of Electrical Insulation at very Low Frequencies', *ibid.*, Paper No. 5.
(19) Webb, J. K., and Wood, H. B.: 'The Precise Measurement of Capacitance', *Proceedings I.E.E.*, Monograph No. 100 M, May, 1954 (**102** C, p. 3).
(20) Ward, W. H., Oliver, M. H., and Fray, S. J.: 'H.F. Resistance Standards and their use in the Calibration of an Admittance Bridge up to 60 Mc/s', *ibid.*, Paper No. 1017 R, November, 1950 (**97**, Part III, p. 438).
(21) Rayner, G. H.: 'A Selective Detector Amplifier for 10–10 000 c/s', *Journal of Scientific Instruments*, 1953, **30**, p. 17.
(22) Arnold, A. H. M.: 'Measurement of Power at Audio Frequencies, with special reference to the Equipment at the N.P.L. for Calibrating Wattmeters', National Physical Laboratory Symposium on Precision Electrical Measurements 1954, H.M. Stationery Office, Paper No. 16.
(23) Arnold, A. H. M.: 'Alternating-Current-Instrument Testing Equipment', *Proceedings I.E.E.*, Paper No. 1532 M, April, 1954 (**101**, Part II, p. 121).
(24) Hermach, F. L.: 'Thermal Convertors as A.C./D.C. Transfer Standards for Current and Voltage Measurements at Audio Frequencies', *Journal of Research of the National Bureau of Standards*, 1952, **48**, p. 121.
(25) Hermach, F. L., and Williams, E. S.: 'Multirange Audio-Frequency Thermocouple Instruments of High Accuracy', *ibid.*, 1954, **52**, p. 227.
(26) Widdis, F. C.: 'The Indirectly Heated Thermistor as a Precise A.C.–D.C. Transfer Device', *Proceedings I.E.E.*, Paper No. 1989 M, February, 1956 (**103** B, p. 693).
(27) Hill, J. J.: 'A Precision Thermo-Electric Wattmeter for Power and Audio Frequencies', *ibid.*, Paper No. 2461 M, January, 1958 (**105** B, p. 61).
(28) Arnold, A. H. M., and Hill, J. J.: 'An Audio-Frequency Dynamometer Wattmeter', *ibid.*, Monograph No. 171 M, March, 1956 (**103** C, p. 325).
(29) Wilkins, F. J.: 'An Electronic Supply for use in the Calibration of Instruments', *ibid.*, Paper No. 1937 M, January, 1956 (**103** B, p. 83).
(30) Smith, W. E., and Clothier, W. K.: 'Determination of the D.C./A.C. Transfer Error of an Electrostatic Voltmeter', *ibid.*, Paper No. 1686 M, October, 1954 (**101**, Part II, p. 465).
(31) Patchett, G. N.: 'A New Precision A.C. Voltage Stabilizer', *ibid.*, Paper No. 929 M, January, 1950 (**97**, Part II, p. 529).
(32) Menzies, A. C.: 'Radiation Measuring Instruments for the Infra-Red to Ultra-Violet Waveband', *ibid.*, Paper No. 1145 M, May, 1951 (**98**, Part II, p. 771).
(33) Taylor, D.: 'Radioactivity Surveying and Monitoring Instruments', *Journal of Scientific Instruments*, 1950, **27**, p. 81.
(34) Birks, J. B.: 'Scintillation Counters' (Pergamon Press, 1953).
(35) Taylor, D., and Abson, W.: 'Measuring Instruments for X-Rays to Cosmic Rays', *Proceedings I.E.E.*, Paper No. 1148 M, May, 1951 (**98**, Part II, p. 760).
(36) Gillespie, A. B.: 'The Control and Instrumentation of a Nuclear Reactor', *ibid.*, Paper No. 2058 M, March, 1956 (**103** B, p. 564).
(37) Curran, S. C., and Craggs, J. D.: 'Counting Tubes' (Butterworth, 1949).
(38) Shaw, E. N.: 'The Alpha Gauge', *Journal of the British Institution of Radio Engineers*, September, 1954, **14**, p. 414.
(39) Syke, G.: 'A Gamma Ray Thickness Gauge for Hot Steel Strips and Tubes', *ibid.*, p. 419.
(40) Symposium of Papers on Radiation Monitoring Apparatus, *Proceedings I.E.E.*, 1951, **98**, Part II, p. 173.
(41) Belvers, H. G.: 'Three Methods of Measuring Magnetic Fields, (3) Measurement by the Proton Resonance Method', *Philips Technical Review*, 1953, **15**, p. 49.

(42) Hopkins, N. J.: 'A Magnetic Field Strength Meter using the Proton Magnetic Moment', *Review of Scientific Instruments*, 1949, **20**, p. 401.

(43) Essen, L., and Parry, J. V. L.: 'The Caesium Resonator as a Standard of Frequency and Time', *Philosophical Transactions of the Royal Society*, 1957, **250**, p. 45.

(44) Kandiah, K., and Brown, D. E.: 'High Gain D.C. Amplifiers', *Proceedings I.E.E.*, Paper No. 1196 M, November, 1957 (**99**, Part II, p. 314).

(45) Buckerfield, P. S. T.: 'The Parallel-T D.C. Amplifier, a Low-Drift Amplifier with Wide Frequency Response', *ibid.*, Paper No. 1361 M, October, 1952 (**99**, Part II, p. 497).

(46) Summerlin, F. A.: 'The Automatic Compensation of Zero-Drift Errors in Direct-Coupled Feedback Systems', *ibid.*, Paper No. 1088 M, February, 1957 (**98**, Part II, p. 59).

(47) Dauphinee, T. M., and Woods, S. B.: 'Low-Level Thermocouple Amplifier and a Temperature Regulation System', *Review of Scientific Instruments*, 1955, **26**, p. 693.

(48) Atkinson, P. D., and Hemingway, A. V.: 'An Even-Harmonic Magnetic Amplifier and Some Applications to Measurement and Control', *Electronic Engineering*, 1954, **26**, p. 482.

(49) Nielsen, S. O., and Rosenberg, T.: 'A Stable and Sensitive D.C. Amplifier with High Input Resistance', *Journal of Scientific Instruments*, 1954, **31**, p. 401.

(50) Dorsman, C.: 'A pH Meter with a Very High Input Resistance', *Philips Technical Review*, 1942, **7**, p. 24.

(51) Hartog, H. den, and Muller, F. A.: 'Verwijdering van den Ruisch in Thermionische Electrometers met Mechanische Conversie', *Physica*, 1944, **11**, p. 161.

(52) Palevsky, H., Swank, R. K., and Grenchik, R.: 'Design of Dynamic Condenser Electrometers', *Review of Scientific Instruments*, 1947, **18**, p. 298.

(53) Chaplin, G. B. B., and Owens, A. R.: 'A Transistor High-Gain Chopper-Type D.C. Amplifier', *Proceedings I.E.E.*, Paper No. 2442 M, November, 1957 (**105 B**, p. 258).

(54) Connor, F. R.: 'An Ultra-High-Speed Oscillograph', *ibid.*, Paper No. 1931 R, March, 1956 (**103 B**, p. 179).

(55) Bjorkman, J., and Lindberg, L.: 'Development of Trochotrons', *Ericsson Technical Journal*, 1954, **10**, p. 3.

(56) Bacon, R. C., and Pollard, J. R.: 'The Dekatron, A New Cold Cathode Counting Tube', *Electronic Engineering*, 1950, **22**, p. 173.

(57) Acton, J. R.: 'The Single-Pulse Dekatron', *ibid.*, 1952, **24**, p. 48.

(58) Hitchcox, G.: 'The Short Interval Timer', *ibid.*, 1956, **28**, p. 298.

(59) Chesterman, W. D., Glegg, D. R., Peck, G. T., and Meadowcroft, A. J.: 'A New Power Stroboscope for High-Speed Flash Photography', *Proceedings I.E.E.*, Paper No. 1127 U, March, 1951 (**98**, Part II, p. 619).

(60) Winch, G. T.: 'The Measurement of Light and Colour', *ibid.*, Pape No. 822 M, January, 1949 (**96**, Part II, p. 452).

(61) Bett, N., Nock, W., and Morris, G.: 'An Automatic Coulometri Titrimeter', *Analyst*, 1954, **79**, p. 607.

(62) Kolthoff, I. M., and Lingane, J. J.: 'Polarography" (Intersciena Publishers Inc., New York, 1952).

(63) Ferrett, D. J., Milner, G. W. C., Shalgosky, H. I., and Slee, L. J. 'A Comparative Study of Three Recently Developed Polarographs' *Analyst*, 1956, **81**, p. 506.

(64) Barber, C. R.: 'Platinum Resistance Thermometers of Small Dimensions', *Journal of Scientific Instruments*, 1950, **27**, p. 47.

(65) Barber, C. R., Gridley, A., and Hall, J. A.: 'An Improved Construction of the Smith Bridge Type 3', *ibid.*, 1955, **32**, p. 213.

(66) Gautier, M.: 'A Modification of the Smith Bridge, Type 3', *ibid.* **30**, p. 381, 1953.

(67) Parker, R. C., and Marshall, P. R.: 'The Measurement of th Temperature of Sliding Surfaces, with particular reference to Railway Brake Blocks', *Proceedings of The Institution of Mechanica Engineers*, 1948, **158**, p. 209.

(68) Barber, C. R., and Pyatt, E. C.: 'Low-Temperature Radiation Pyrometry using a Lead Sulphide Photo-Conductive Cell', *Nature*, 1950 **165**, p. 691.

(69) Barber, C. R., and Pyatt, E. C.: 'An Optical Pyrometer employin an Image-Converter Tube for use over the Temperature Rang 350–700° C', *Journal of Scientific Instruments*, 1950, **27**, p. 4.

(70) Smollett, M., and Jenkins, J. A.: 'The Lead Sulphide Photo-Conductive Cell', *Electronic Engineering*, 1956, **29**, p. 373.

(71) McLean, J. A.: 'A Method for Constructing Direct Reading Thermistor Thermometers', *Journal of Scientific Instruments*, 1954, **31** p. 455.

(72) Edmundson, D.: 'Electric and Magnetic Measurements in an Electrical Engineering Factory', *Proceedings I.E.E.*, Paper No. 1703 M September, 1954 (**102 B**, p. 427).

(73) Shortt, H. L., and Eisler, P.: 'Printed Foil Electronic Components' *Tele-Tech*, 1955, **14**, p. 404.

(74) Yates, J. G., Lucas, D. H., and Johnston, D. L.: 'Multi-Channe Measurement of Physical Effects by Confluent Pulse Technique with particular reference to the Analysis of Strain', *Proceeding I.E.E.*, 1951, Paper No. 1032 M (**98**, Part II, p. 109).

(75) Corner, W. D., and Hunt, G. H.: 'A Direct Reading Instrument fo the Measurement of Small Displacements', *Journal of Scientifi Instruments*, 1954, **31**, p. 445.

(76) Andrews, H. I.: 'Electrical Weighing', *Proceedings I.E.E.*, Pape No. 894 M, October, 1949 (**97**, Part I, p. 98).

(77) Woodcock, F. J.: 'Some Electrical Methods of Measuring Mechanica Quantities', *ibid.*, Paper 901 M, July, 1950 (**97**, Part I, p. 136).

:21.397.5

The Institution of Electrical Engineers
Paper No. 2540 R
Feb. 1958
©

THE RELATION BETWEEN PICTURE SIZE, VIEWING DISTANCE AND PICTURE QUALITY

With Special Reference to Colour Television and Spot-Wobble Techniques

By L. C. JESTY, B.Sc., Member.

(The paper was first received 1st July, and in revised form 26th November, 1957. It was published in February, 1958, and was read before the SOUTH-WEST SCOTLAND SUB-CENTRE *11th February*, the SOUTH-EAST SCOTLAND SUB-CENTRE *12th February, and the* RADIO AND TELECOMMUNICATION SECTION *19th February*, 1958.)

SUMMARY

The paper describes experiments to determine the preferred viewing distance for a number of different types and sizes of picture, including 405-line and 625-line monochrome television and a 405-line colour television picture, all with varying bandwidths. The criterion 'preferred viewing distance' proved to be surprisingly consistent and repeatable, and the results obtained indicate that, unless some new development such as receiver picture storage improves matters, the present European standards of television are wasteful of bandwidth. Various methods of using this surplus bandwidth are discussed. The investigation also covered the use of spot-wobble on the 405-line television picture and two television scanning innovations—'synchronous spot-wobble' and 'sampled synchronous spot-wobble'.

LIST OF SYMBOLS

V = Chosen viewing distance.
H = Height of picture as viewed.
A = Area of picture as viewed.
V/H = Ratio of viewing distance to picture height—a criterion of picture quality.
f = Video-frequency bandwidth of luminance channel.
N = Number of lines.
σ = Standard deviation of observations (expressed as a percentage of mean).
S = Circle of confusion of projected image.
Y = Projector-lens—screen distance for 'sharp' focus.
X = Distance projector is pulled back from sharply focused position.
ϕ = Angle subtended by incident beam from projector on to sharp image plane.
α = Parameter relating to the design of low-pass filters (from Reference 4).

(1) INTRODUCTION

The present investigation arose directly as a result of the C.C.I.R. Colour Tour (Spring, 1956). On a number of occasions formal demonstrations of colour television were given, with the front row of the audience at about 12 times the picture height from the displays. At this distance it is easily estimated from previous work[1] that something less than 1 Mc/s luminance bandwidth is sufficient to satisfy the average observer. A 405-line N.T.S.C. type colour-television picture was set up with a fully-corrected luminance channel of 0·75 Mc/s, and a team of observers were invited to choose where they should sit to view this. Their average viewing distance was 13·2 times the picture height, and indeed, at this viewing distance, the picture was quite acceptable. For a 2·25 Mc/s luminance channel they chose 7·7 times the picture height. The behaviour of the observer team in carrying out this simple task was so consistent and repeatable that it was decided to extend the investigation. Colour transparencies and films were projected at various picture sizes and degrees of 'sharpness'. 405-line and 625-line closed-circuit monochrome television systems were tested with various video-frequency bandwidths. The 405-line experiments included variations of spot-wobble and synchronous spot-wobble,[1,2] which produced considerable improvement in picture quality.

Mr. Jesty is with Sylvania-Thorn, Ltd.

(2) OPTICAL-PROJECTION EXPERIMENTS

The object of the optical-projection experiments was to obtain general viewer reaction, in terms of chosen viewing distance, to pictures of different sizes and degrees of sharpness. Four matt-white 4 : 3 viewing screens were used, 9, 13, 18 and 26 in high, i.e. a ratio of approximately 2 : 1 in area between each size. Brightness was kept constant between the different picture sizes by placing suitable neutral filters over the projector lens. Picture-highlight brightness was kept in the region of 20 ft-lamberts. The experiments were carried out in a darkened room approximately 26 ft × 12 ft, and with some low ambient lighting which did not spoil the contrast in the projected picture. The observers were dealt with one at a time and the tests were carried out in a reasonably random order. The observers were not told anything about the test conditions; they were merely asked to take a chair and sit where they would most like to view the picture. They appeared to have no difficulty in carrying out this task in a few moments, and the distance from their eyes to the screen was then measured. Relevant technical details are given in Section 8.

Two basically different types of experiment were carried out. The method of the first series is shown in Fig. 1 and the results are given in Fig. 2. The basic principle of the second series is shown in Fig. 3 and the results in Figs. 4 and 5. In all these

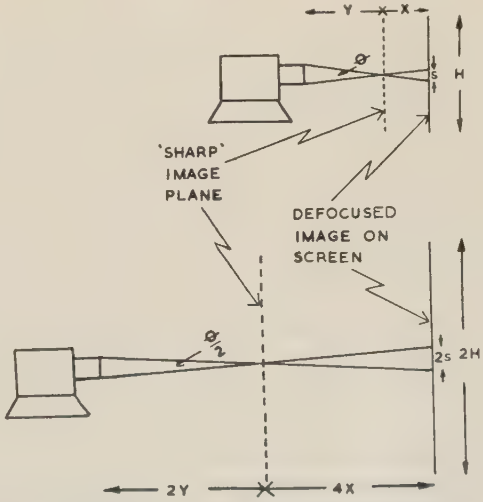


Fig. 1.—Optical projection of pictures of equal sharpness and different sizes by proportionate defocusing.

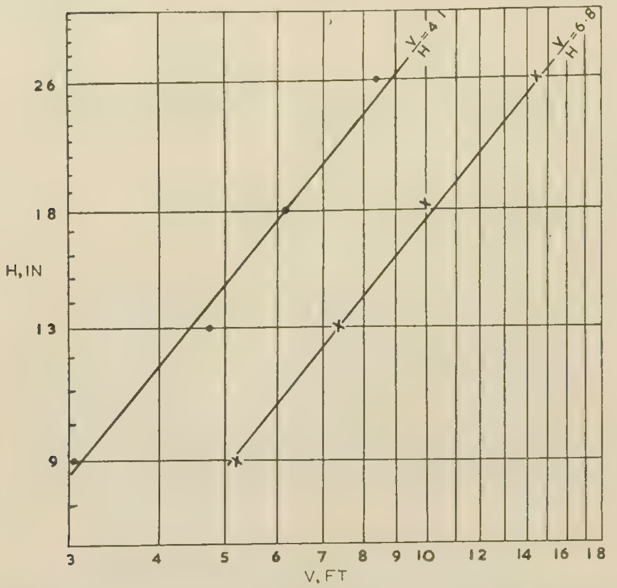


Fig. 2.—Viewing distance chosen for optically-projected pictures of different size and equal sharpness.

● 'Sharp' projection.
× 'Defocused' projection.

Each point is the mean of nine observers viewing four 35 mm colour transparencies, i.e. equivalent to 36 observations.

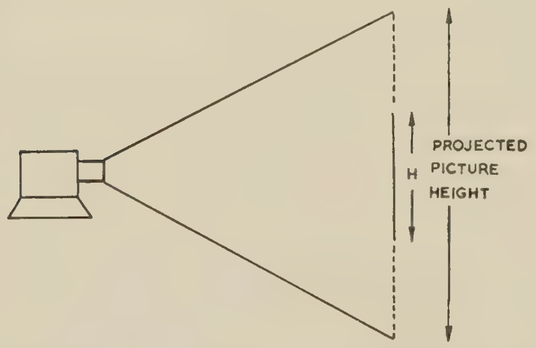


Fig. 3.—Optical projection of pictures of known relative sharpness, by displaying measured parts of the image.

experiments nine observers were used and four colour transparencies were projected, so that each point plotted in Figs. 2, 4 and 5 represents the average of 36 individual observations.

In the first series of experiments the colour transparencies were projected sharply focused on to the four screens. They were then projected defocused by a given amount, using the method shown in Fig. 1. The projector was sharply focused on the screen and then pulled back a predetermined amount. A rough estimate was made of the distance which the projector had to be pulled back to give a picture comparable with the above 405-line colour-television picture. Various methods were tried, including calculation of the circle of confusion and the use of resolution charts. The most satisfactory method appeared to be direct comparison of pairs of matched colour transparencies, one of which was projected and defocused by the method of Fig. 1 until it resembled the other transmitted via the colour-television system. Results obtained with the viewer test are given in Fig. 2, and they show quite clearly that, for constant picture sharpness, the viewers moved proportionately farther away as the picture size was increased. For the sharply focused colour transparencies they chose to view at 4·1 times the picture height,

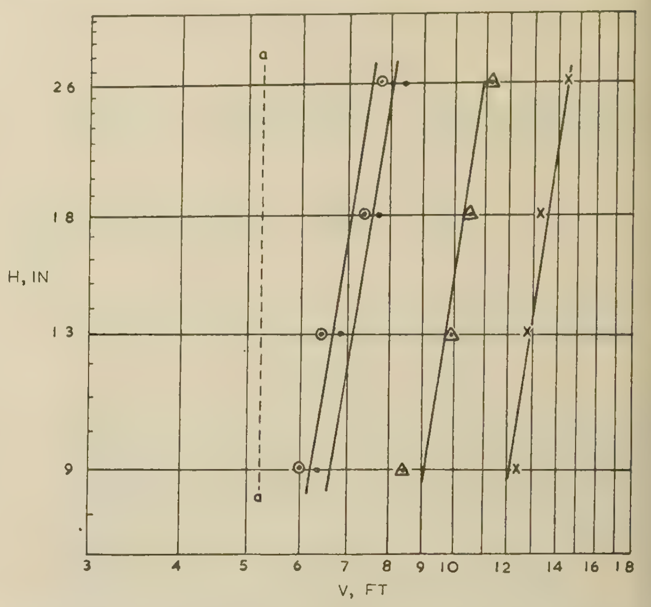


Fig. 4.—Viewing distance chosen for optically projected pictures of different size, sharpness being proportional to picture height.

● 'Sharp' image projected 26 in high.
× 'Defocused' image projected 26 in high.
△ 'Sharp' image projected 55 in high (one-quarter normal brightness).
⊙ 'Sharp' image projected 26 in high (one-quarter normal brightness).

Each point is the mean of nine observers viewing four pictures, i.e. equivalent to 36 observations.
Equation of $a \ldots a$ is V = constant.

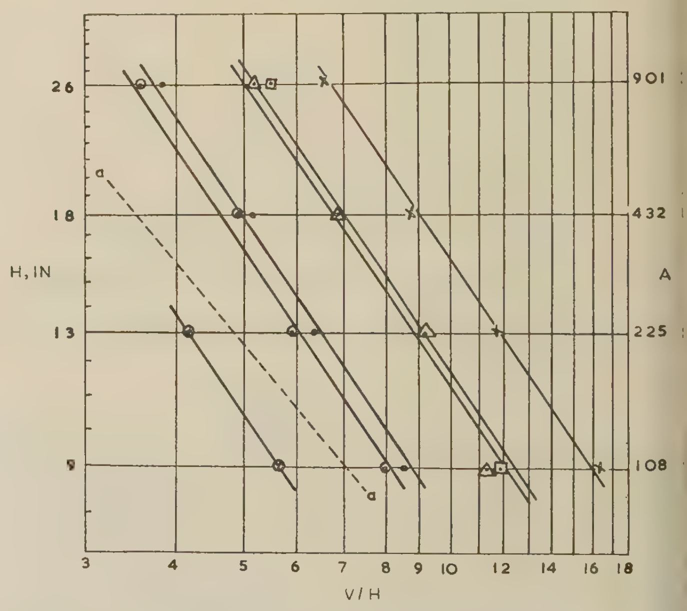


Fig. 5.—Ratio of viewing distance to picture height chosen for optically projected pictures of different size, sharpness being proportional to picture height (i.e. 'bandwidth' proportional to picture area).

● 'Sharp' image projected 26 in high.
× 'Defocused' image projected 26 in high.
△ 'Sharp' image projected 55 in high (one-quarter normal brightness).
⊙ 'Sharp' image projected 26 in high (one-quarter normal brightness).
⊗ 'Sharp' image projected 13 in high (four times the normal brightness).

Each point is the mean of nine observers viewing four 35 mm colour transparencies, i.e. equivalent to 36 observations.

⊡ 16mm original colour film projected 26 in high (nine observers).
Slope of lines through experimental points is $H(V/H)^{1\cdot 23}$ = constant.
Equation of $a \ldots a$ is $H(V/H)$ = constant.
Picture area $A = 4/3H^2$ square inches.

irrespective of size. For the defocused pictures they chose 8·8 times the picture height (compared with 7·7 for the 405-line colour-television picture). Both these experiments show the constancy of the ratio of viewing distance to viewing height, V/H, for a given degree of picture sharpness, and the difference between the two experiments shows quite clearly that viewers will make good use of an increase in sharpness by sitting closer to the picture.

The second series of experiments were aimed at obtaining the correlation between the viewing distance and the picture sharpness. The results obtained in the first experiment showed a very consistent behaviour towards pictures with two different degrees of sharpness, but the method did not lend itself to a very accurate assessment of the relative sharpness of the pictures used. The method of Fig. 3 was therefore devised in an attempt to obtain accurate relative assessment of picture sharpness and correlate it with observer reaction. Fig. 3 shows a screen of height smaller than the image projected, placed in the image plane so that the observer only sees part of the picture. It is assumed that the sharpness of the projected image is the same all over, and then the sharpness of the picture as viewed will be directly proportional to the height H of the picture. By this procedure different amounts of the actual picture were displayed, exactly on the lines of Fig. 18 in an earlier paper.[3] The observers were, in fact, being offered a picture of constant sharpness per unit area and of increasing size. For constant acuity of vision, therefore, they should choose to sit at a constant viewing distance for all picture sizes. This means that the experimental results when plotted in Fig. 4 would be expected to have the slope shown by the dotted line $a \ldots a$. In fact, as can be seen in every case, they were slightly inclined to this, the observers moving quite consistently a little farther away from the larger pictures despite the constant sharpness per unit area. Three different projected image sizes were used in these experiments, 13, 26 and 55 in high. The defocused condition of Fig. 1 was also used, projected 26 in high, and changes in brightness were also investigated. Some original 16 mm colour film was projected 26 in high. The data are plotted in terms of V/H in Fig. 5, from which it can be seen that all the experiments produce practically a straight-line relationship on the log/log scales, the power of the law connecting H and V/H being $-1 \cdot 23$. The line $a \ldots a$ from Fig. 4 is shown replotted. This, of course, has a -1 power relationship. These powers become $-2 \cdot 46$ and -2 if the relation between V/H and picture area is taken. (Picture area is shown plotted on the right-hand side of Fig. 5.) In equating this diagram to the television experiments to follow, picture area can be taken as being directly proportional to video-frequency bandwidth.

As expected from the earlier work[1], a change in brightness over the range 5–80 ft-lamberts has only a second-order effect on apparent picture sharpness, measured in terms of V/H. From Fig. 5 an increase of 4 : 1 in brightness would appear to produce an increase of about 10% in V/H. The previous work showed an increase in acuity of about 22% in the observers for an increase in brightness of 10 : 1 in the range 1–100 ft-lamberts highlight brightness.

One discrepancy in these experiments requiring further investigation is the fact that, although the 55 and the 26 in high conditions behave consistently within themselves and give lines of sensibly equal slope, their relative positions in the diagram are not as would be expected. From the previous experiment of Fig. 2 it would be expected that the image 55 in high viewed on a screen 26 in high would give sensibly the same value of V/H as the 26 in image of $\frac{1}{4}$ brightness viewed on the 13 in screen. The values of V/H for these two conditions are almost exactly 4 and 6, respectively. On the other hand, the image projected 55 in high and viewed on the 13 and 9 in screens is in very good agreement with the 26 in data, making due allowance for the brightness difference.

Only two points are available for the 16 mm colour projection, and taken by themselves they indicate a higher slope for the line passing through them.

(3) TELEVISION EXPERIMENTS

The same procedure of nine observers viewing four pictures was carried out with 405- and 625-line closed-circuit television systems. A 35 mm monochrome slide scanner for 405- and 625-line scanning with a 21 in display monitor was used and a 35 mm colour slide scanner for a 405-line version of the N.T.S.C. system already referred to in the Introduction. The picture in this case was displayed on a 21 in tri-colour tube operating from a receiver modified to British standards. The signal from the flying-spot scanner was fed in on a 45 Mc/s r.f. carrier. The viewing monitors were set up in the room used for the optical experiments with the picture in approximately the same position. The same four colour slides were used in all the tests.

An experimental 405-line $4\frac{1}{2}$ in image-orthicon camera chain was also tested using a 17 in rack-mounted monitor in the laboratory. This had receiver spot-wobble, synchronous spot-wobble and sampled synchronous spot-wobble available,[2] and tests were made under all the conditions of operation. In making their observations on this equipment the viewers had to stand in the laboratory. Four pictures were used, but, in this case, they were 10 in × 8 in black-and-white transparencies illuminated from behind. The type of subject-matter was very similar to that used in the optical and slide scanner tests.

Three fully-corrected low-pass filters were used to limit the video-frequency bandwidth of the television systems during certain of the tests. These filters[4] had bandwidths of 3·0, 1·5 and 0·75 Mc/s. Details of their characteristics are given in Table 1. The maximum bandwidths of the monochrome flying-spot and $4\frac{1}{2}$ in image-orthicon equipment were greater than 7 Mc/s. The luminance channel of the flying-spot colour equipment was limited to 2·25 Mc/s, i.e. cut-off at sub-carrier frequency.

Table 1

CHARACTERISTICS OF LOW-PASS FILTERS

Filter bandwidth	6dB attenuation point	First zero
Mc/s	Mc/s	Mc/s
3·00	3·45	3·75
1·50	1·73	1·90
0·75	0·87	0·95

In all these television tests great care was exercised to ensure that the receiver displays were interlacing properly. Further relevant details of the tests are given in Section 8.

The results of all the tests with the monochrome equipment using simple 2 : 1 interlaced scanning at the four different bandwidths are given in Fig. 6. It should be noted that the 0·75 Mc/s point for the 625-line system gives $V/H > 18$, which is greater than the length of the room available. Attention should be paid mainly to the individual experimental points themselves and not to the curves which have been drawn through them. As in the optical experiments, each point represents the average of nine observers looking at four pictures, and it can be seen that the repetition accuracy is again quite high. Some of the repeat experiments were carried out with a lapse of several weeks. There is also good agreement in the 405-line tests between the flying-

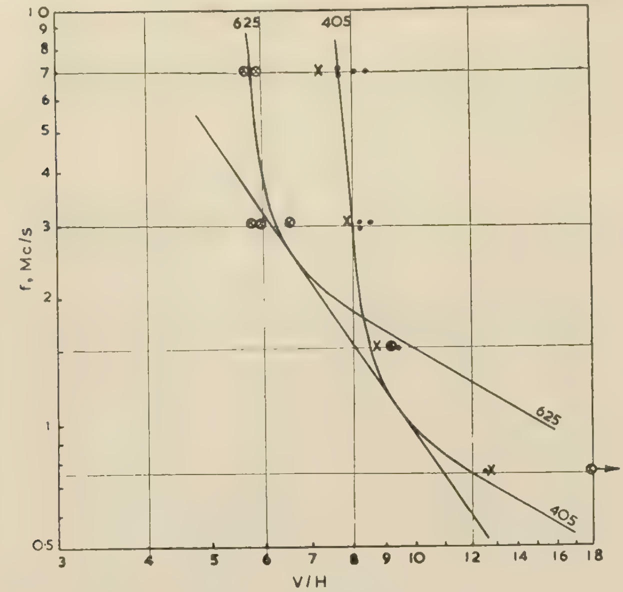


Fig. 6.—Ratio of viewing distance to picture height chosen for 405- and 625-line monochrome television pictures of different bandwidths.

● 405 line $4\frac{1}{2}$ in I.O. camera.
× 405 line flying-spot 35 mm slide scanner.
⊗ 625 line flying-spot 35 mm slide scanner.

Each point is the mean of nine observers viewing four pictures—36 observations. Equation of common tangent line is $(V/H)^{2\times 1\cdot 23}$ = const = 260.

spot scanner and the $4\frac{1}{2}$ in image orthicon equipment, although the flying-spot equipment is slightly superior.

Unlike the optical experiments, the television tests do not result in straight lines of constant slope on the log/log plot. The work of M. W. Baldwin[5] on picture-reproducing systems, in which the ratio of vertical to horizontal resolution was varied, showed a marked maximum for the condition where resolution is equal in the two directions. Performance fell off fairly symmetrically on either side of this optimum, as the ratio of the resolving power in the two directions was increased above it or reduced below it. This is shown quite clearly in Fig. 5 of his paper, which is reproduced in Fig. 7. In the present series of experiments it would be expected, therefore, that, at any given bandwidth, there would be an optimum number of lines giving a minimum value of V/H.

Although we must not attempt to deduce too much from the present experiments, it is quite obvious that the 405- and 625-line data do give two curves of the expected form, and that the tangent to these curves, representing the locus of the optimum combination of viewing distance and bandwidth for each system, has a slope very similar to that of the lines for the optical experiments plotted in Fig. 5. In making this comparison the area scale on the right-hand side of Fig. 5 is, of course, taken as representative of the bandwidth. Deliberately taking a tangent slope of $-2\cdot 46$, the data in Fig. 6 have been replotted in Fig. 8 so that the 405- and 625-line points are 'fitted' on to a common curve. In addition to 'sliding' the data down the tangent line to make them overlap, it is also necessary to decide the exact amount by which they shall be 'slid'. It will be remembered that the basic principle behind the experiments of Fig. 5 was that, as picture size was increased, the resolution per unit area was kept constant. The assumption made in plotting Fig. 8 is that the higher-definition 625-line system working at its optimum bandwidth ('tangent point') is used to produce a proportionately larger picture at the same viewing distance than the 405-line system working at its optimum bandwidth. This means that at the two tangent points in Fig. 6 the ratio of the two horizontal 'sharpnesses' and that of the two vertical 'sharpnesses' are taken to be equal, and equal to 625/405. On the bandwidth scale, the data for the two systems have been 'slid' in the ratio $(625/405)^2$.

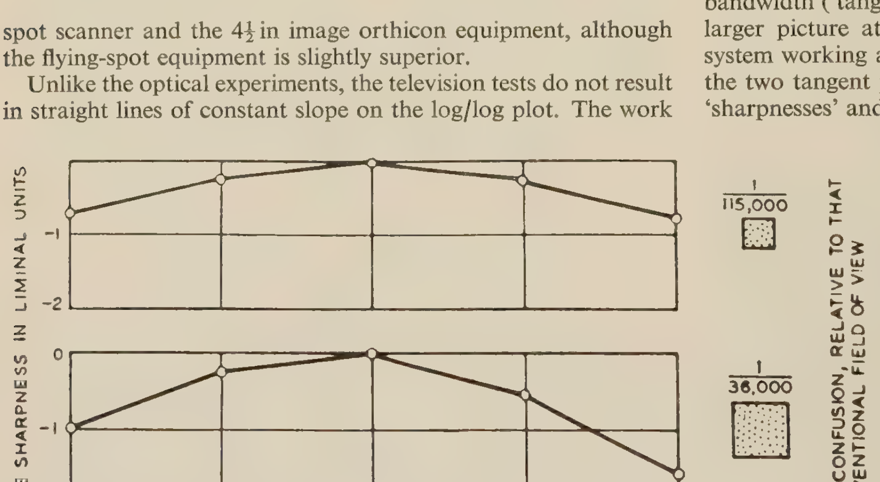


Fig. 7.—Sharpness of small-sized motion pictures as a function of the relative values of horizontal and vertical resolutions (reproduced from Reference 5).

The conventional field of view is a rectangle whose height is one-quarter the viewing distance and whose width is four-thirds the height. Each point represents 150 observations at a viewing distance of 30 in.

Fig. 8 shows a very satisfactory superposition of the two sets of data and indicates that individual television systems could be represented by a family of curves suitably positioned along the tangent line in Fig. 6. The two curves shown in Fig. 6 are, in fact, the average curve of Fig. 8 plotted back in its correct positions for the two systems.

The probable error of the data is important and has been calculated as follows. The means of the readings for the four pictures were averaged for each observer. The standard deviation of these nine results was then determined for each experiment (each point in Figs. 6 and 8). These standard deviations, expressed as a percentage of their associated mean, were then averaged throughout the whole series of tests. This gave a value for σ of $\pm 14\%$ for any one observer. For nine observers, therefore, the variability will be of the order of one-third of this. The shaded area in Fig. 8 is for $\pm\frac{2}{3}\sigma$ limits on the mean. It would be expected, therefore, that approximately 1 in 20 of the experimental points would fall outside the shaded area. There are, in fact, 20 points plotted in Fig. 8, and one does just lie outside the top limit line. The indication is that the results are self-consistent and that the assumptions made in deriving Fig. 8 from the data in Fig. 6 have a reasonably

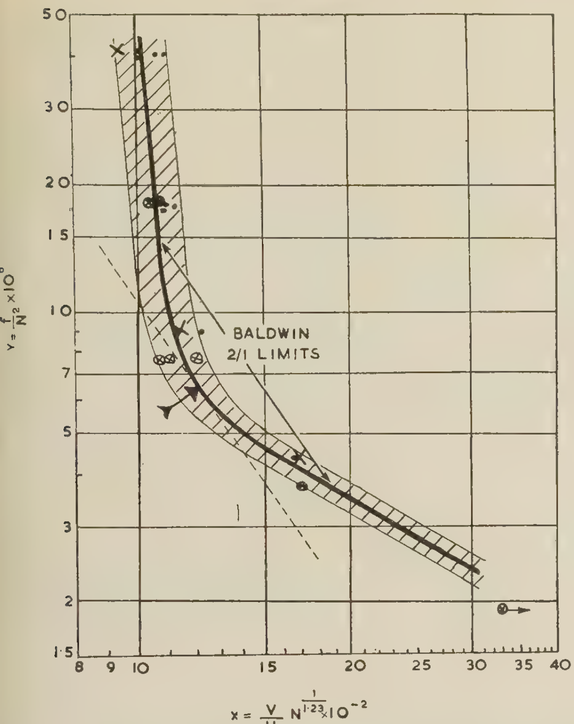


Fig. 8.—Superimposition of 405- and 625-line experimental data.
The arrowhead indicates the optimum choice for minimum bandwidth.
The equation of the common tangent line is

$$X^{2 \times 1 \cdot 23} Y = \text{constant} = 3100$$
$$(\text{or } f(V/H)^{2 \times 1 \cdot 23} = 260)$$

The shaded area indicates 0·95 probability limits.

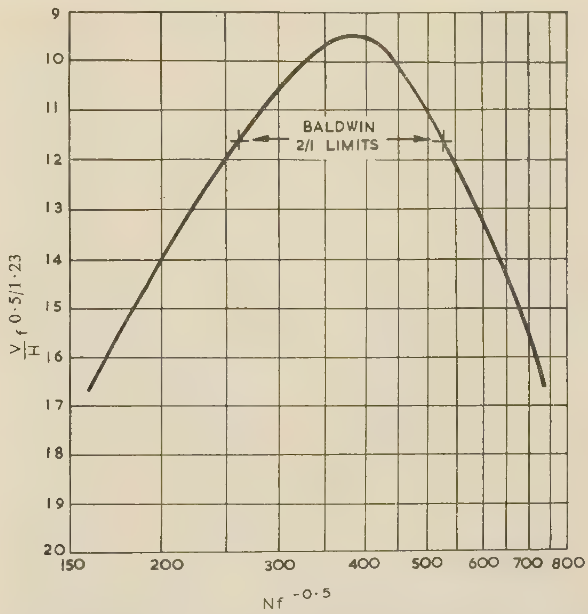


Fig. 9.—Efficiency of a television system of a given bandwidth for different numbers of lines, expressed in terms of V/H.

$$\frac{V}{H} f^{0 \cdot 5/1 \cdot 23} = 10^{2-(3/1 \cdot 23)} X Y^{0 \cdot 5/1 \cdot 23}$$

$$N f^{-0 \cdot 5} = 10^3 Y^{-0 \cdot 5}$$

where X and Y are the abscissae and ordinates of Fig. 8.

high probability of accuracy. (Expressing the standard deviation of the optical experiments Figs. 2, 4 and 5 as a percentage of their means, and averaging in the same way gave $\sigma = \pm 19\%$ for one observer looking at four pictures.)

As already stated, Fig. 8 can be made to represent a family of curves lying along an optimum tangent line, each curve being for a given number of lines. Two such curves are shown in Fig. 6. The point of contact of each individual curve with the tangent represents the optimum bandwidth for that number of lines. Any other choice of number of lines for that bandwidth will result in an inferior performance in terms of closeness of viewing distance (measured in terms of V/H).

The deductions from Fig. 6 are rather surprising. The optimum bandwidth for a 405-line system is just over 1 Mc/s, and for a 625-line system it is about $2\frac{1}{2}$ Mc/s. As will be seen from the curves, there is still considerable advantage to the viewer in increasing the bandwidth above these values, but not nearly so much as would be gained by increasing the number of lines as well.

The efficiency of any given choice of number of lines and bandwidths can now be judged on an absolute basis from Figs. 6 and 8, by taking the ratio of V/H for the system as chosen to that for the point where the same bandwidth intersects the tangent line. For example, the British 405-line 3 Mc/s standard gives a value for V/H of 8. The tangent line intercepts 3 Mc/s at $V/H = 6 \cdot 2$. On this basis, the system has an overall efficiency of 78% in terms of V/H. The same assessment can be made in terms of bandwidths, in which case, in the example quoted, the same viewing distance can be achieved at the tangent line with a bandwidth of 1·5 Mc/s, which makes this system 50% efficient in terms of bandwidth.

Using the ratio of V/H as the measure of system efficiency, Fig. 9 has been calculated from the data in Fig. 8. This shows, for any chosen bandwidth, the relationship between V/H and the number of lines. It should be noted that the V/H scale has been inverted in order to produce an apparent maximum to the curve. This corresponds to the optimum tangent point shown by the arrowhead in Fig. 8. Baldwin's work, already referred to,[5] showed that, for a given bandwidth, the ratio of the resolving power in the vertical and horizontal directions could change to 2 : 1 in either direction on opposite sides of the optimum (Fig. 7) with little degradation of picture quality. These limits are shown in Fig. 9, and indicate that a degradation in picture quality which produced an increase in V/H of about 20% from the optimum (tangent line) condition would be just acceptable. These limits have been referred back on to Fig. 8.

In order to prevent the author from drawing too many rash conclusions from his curve-fitting exercises, Fig. 10 has been constructed from the experimental points given in Fig. 6. Fig. 10 is derived on exactly the same basis as Fig. 8, but using a tangent slope corresponding to the line $a \ldots a$ in Fig. 5. Again the experimental points lie nicely inside the shaded area!

(4) SPECIAL TELEVISION EXPERIMENTS

The flying-spot colour-television equipment, already referred to, was tested at the full bandwidth of the luminance channel, i.e. 2·25 Mc/s, and with the luminance channel restricted to 0·75 Mc/s. In Fig. 11 the curves, but not the experimental points from Fig. 6, are plotted. It will be seen that the two points for the colour experiments are well within the expected limits from the 405-line monochrome curve.

During the recent colour-television field tests carried out in this country, 83 questionaries, filled in by viewers participating in the tests with standard monochrome receivers, passed through the author's hands. The results showed no significant difference in their values of V/H for values of H ranging from 6 to $13\frac{1}{2}$ in (9 to 21 in receivers). The average value of V/H for the 83 questionaries was 8·8 and this point is shown plotted on Fig. 11 for a bandwidth of 2·5 Mc/s. Again this result lies within the expected limits from the 405-line data obtained in the above

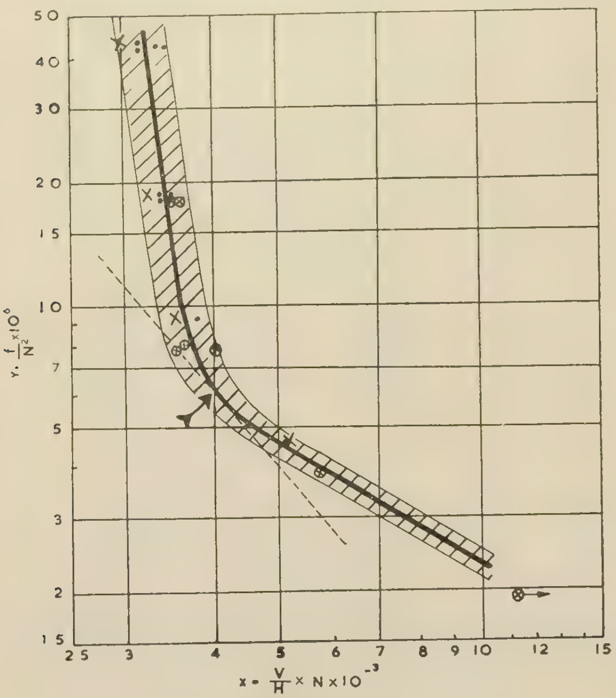


Fig. 10.—Superimposition of 405- and 625-line experimental data.

The arrowhead indicates the optimum choice for minimum bandwidth.
The equation of the common tangent line is

$$X^2Y = \text{constant} = 99$$
$$(\text{or } f(V/H)^2 = 99)$$

The shaded area indicates 0·95 probability limits.

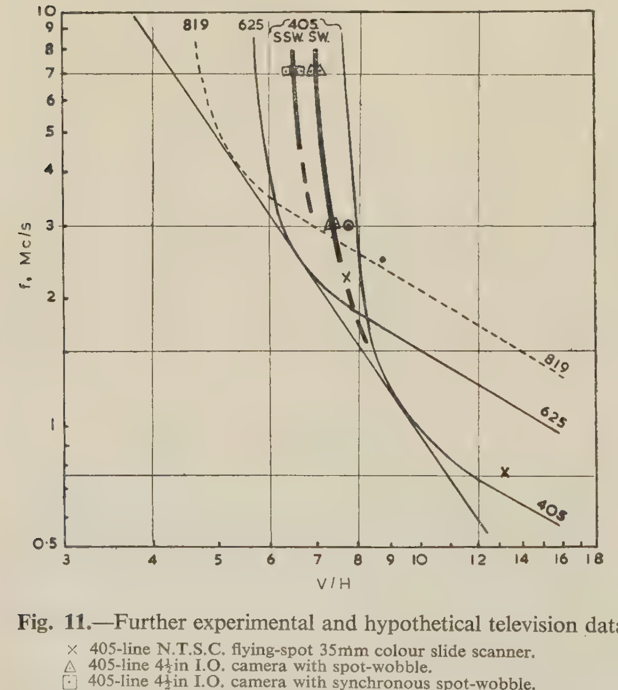


Fig. 11.—Further experimental and hypothetical television data.

× 405-line N.T.S.C. flying-spot 35mm colour slide scanner.
△ 405-line $4\frac{1}{2}$in I.O. camera with spot-wobble.
⊡ 405-line $4\frac{1}{2}$in I.O. camera with synchronous spot-wobble.
○ 405-line $4\frac{1}{2}$in I.O. camera with sampled synchronous spot-wobble.
● 405-line mean of 83 home viewing tests.

laboratory experiments. The value of σ for the 83 results was $\pm 26\%$.

As already mentioned, the $4\frac{1}{2}$ in image-orthicon camera chain had various forms of spot-wobble available. A full description of this equipment and the spot-wobble techniques used is given in the paper by Sarson and Stock,[2] but for clarity a very brief résumé will be given, and Figs. 12, 13, 14, 15 and 16 are reproduced from this paper. Fig. 12 shows a part of a test chart reproduced by a normal 405-line 2 : 1 interlaced system. Fig. 13 shows part of a single field of $202\frac{1}{2}$ lines with receiver spot-wobble applied in the manner described in an earlier paper.[1] By this means the

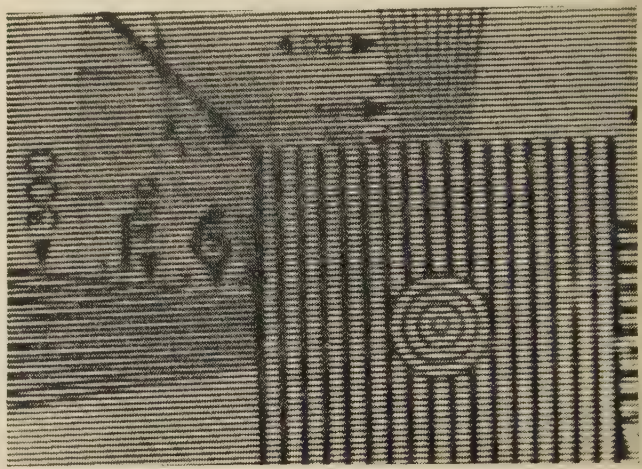


Fig. 12.—Two fields interlaced (reproduced from Reference 2).

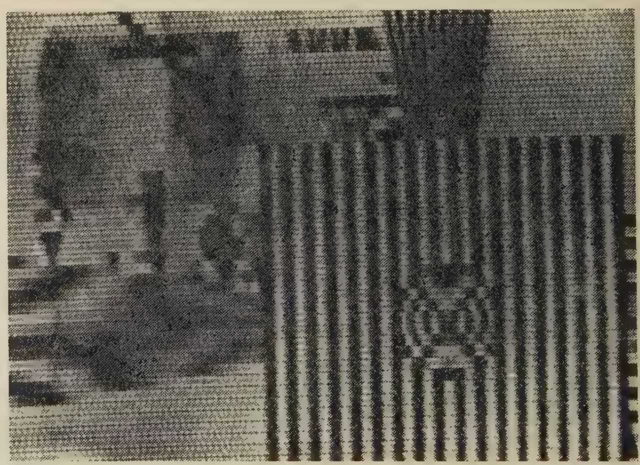


Fig. 13.—Single field with receiver spot-wobble (reproduced from Reference 2).

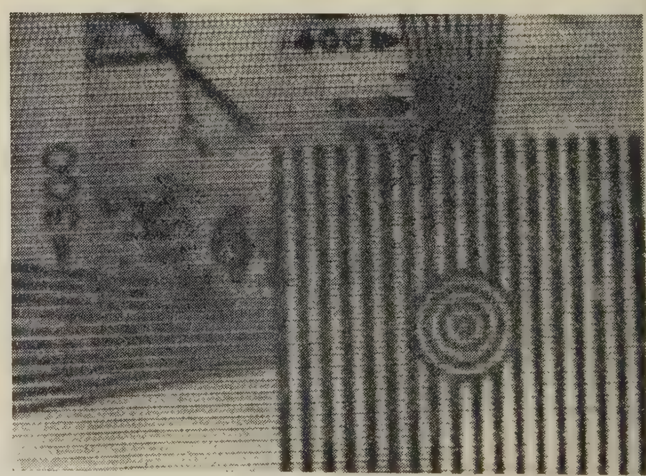


Fig. 14.—Single field with synchronous spot-wobble (reproduced from Reference 2).

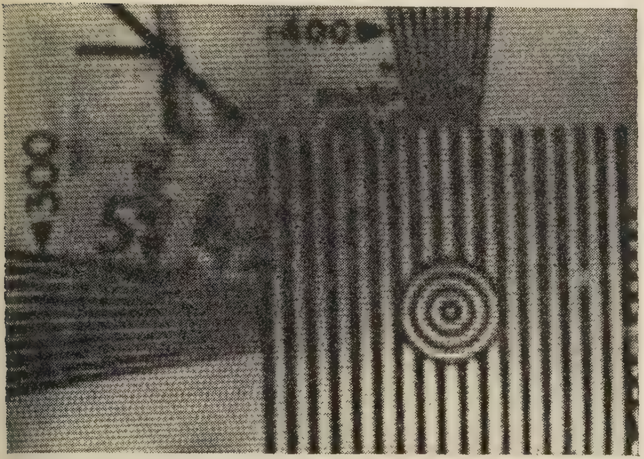


Fig. 15.—Two fields interlaced with synchronous spot-wobble (reproduced from Reference 2).

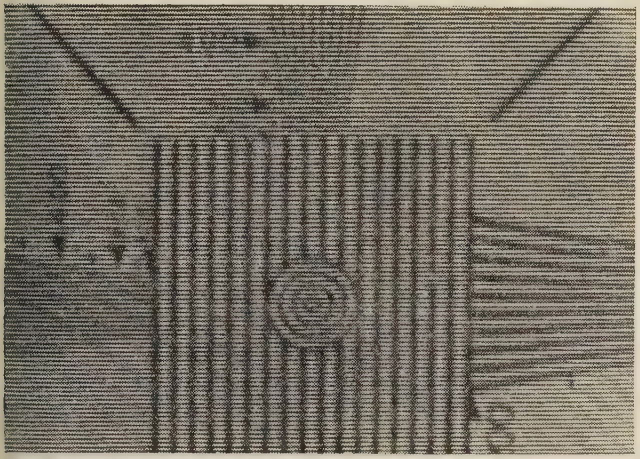


Fig. 16.—Two fields interlaced. Signal sampled from synchronous spot-wobble signal (reproduced from Reference 2).

visibility of the line structure on the viewing screen is reduced but with some slight loss of vertical resolution. Fig. 14 shows the same single field but with spot-wobble applied synchronously to the transmitter camera and the receiver display. By the use of such synchronous spot-wobble, in conjunction with increased video-frequency bandwidth, not only can the loss in vertical resolution be overcome, but vertical resolution can be increased as shown. Fig. 15 shows two fields interlaced from Fig. 14, and is representative of the picture as seen by the viewer on the monitor. This should be compared with Fig. 12, which is the original normal interlaced picture. The increase in resolution in the horizontal wedge can be clearly seen. The beating patterns which are present in Fig. 12 have also practically disappeared.

The synchronous-spot-wobble technique is a method of increasing the vertical resolution of any given standard of television. It has 'compatibility' of scanning standards so that a picture radiated with the higher-definition synchronous spot-wobble could still be received on a normal receiver, but with some small smearing of the vertical resolution. A preferred method of making such a system fully compatible with the existing system would be to radiate the higher-definition picture on new channels, and to obtain the standard television signal from it by sampling in synchronism with the spot-wobble at constant phase. Such a synchronously-sampled spot-wobble signal is shown in Fig. 16. Comparing this with Fig. 12, it will be seen that the vertical wedge is now showing some beating patterns owing to the horizontal sampling involved.

The synchronous-spot-wobble idea is due in its original conception to Blumlein,[6] but the synchronous sampling feature is of more recent origin.[7] Such a system could be used, for example, to produce a high-definition colour-television picture in the u.h.f. bands, the normal standard of definition being sampled out from the luminance channel and radiated in monochrome from existing stations on Bands I and III.

The results of some limited tests carried out on the $4\frac{1}{2}$ in image-orthicon chain with the application of spot-wobble, synchronous spot-wobble and sampled synchronous spot-wobble are shown in Fig. 11. The results of the earlier work[1] are confirmed, namely that receiver spot-wobble alone produces a significant improvement in picture quality. Synchronous spot-wobble produces a further improvement, and the sampled synchronous spot-wobble was apparently judged by the observer team to be not significantly different from the normal 405-line picture.

So far, only preliminary results have been obtained with the various spot-wobble techniques, but nevertheless the results are encouraging. For bandwidths of 3 Mc/s and upwards on a 405-line system, they still do not compare with 625-line scanning, but further investigation into spot-wobble waveform, frequency, etc., could lead to worthwhile improvements. There was some evidence from these experiments that the brightness characteristic of the camera tube was related to the choice of scanning standards. This is in itself an interesting matter for further investigation.

In Fig. 11 the dotted curve shows a hypothetical 819-line system, which is calculated from Fig. 8. The optimum bandwidth appears to be between 4 and 5 Mc/s.

Before proceeding further it is necessary to draw attention to the interpretation of the term 'bandwidth'. This will no doubt be more profitably discussed in detail elsewhere and in the light of further investigations, particularly in relation to the most efficient method of occupying a 'channel'. In the present experiments a 'bandwidth of 3 Mc/s' is specifically related, of course, to the appropriate filter characteristic given in Table 1, and defines the 'flat' part of the characteristic. In this case, therefore, the attenuation characteristic of the filter would not permit this signal to be accommodated in the standard 405-line 5 Mc/s channel, which has a $3\cdot5$ Mc/s sound–vision carrier spacing. The filter would be too wide by at least $0\cdot25$ Mc/s. By a strict interpretation of the data given in the Table, therefore, the frequency scales of Figs. 6 and 8–11 are too low, in terms of television channel allocations. To offset this, however, it seems likely that more efficient filters could be used in any practical system without detriment to picture quality. In any case it is hoped that, in the following discussion, the present results have been interpreted with reserve.

(5) CONCLUSIONS

The most important conclusion from the experiments is that every existing television system working at 25 pictures, 50 fields/sec, has a considerable excess of horizontal resolution which is not being used by the viewer. This conclusion is supported by subjective observer data which are extraordinarily self-consistent and repeatable. Whilst the single criterion used—namely viewing distance—may be criticized as being insufficient to embrace all the qualities looked for in different systems of picture reproduction, nevertheless it does appear to be directly related to the qualities which are important when choosing television scanning standards.

The results are a pointer to the correct assessment of the effectiveness of any signal compression system which may be proposed, for example, from a study of information theory.

The near-optimum bandwidth of $1\frac{1}{2}$ Mc/s for a 405-line system or 3 Mc/s for a 625-line system without spot-wobble give a Kell factor* of 0·42. Baldwin's figure of 0·7 for a sequential scan[5] suggests that a figure in the region of 0·35 should apply to a completely paired interlaced scan. The present results therefore show only very slight advantage to interlacing. Previous work[1] and other similar investigations have shown the ability to resolve the detail in a resolution chart to practically double this amount, suggesting that some other factor must be operating to affect viewers' judgment. The fact that receiver spot-wobble has such a marked effect may be the clue to this. There is always some small vertical movement present in the average programme material, and the $202\frac{1}{2}$-line structure in a single field of a 405-line system immediately becomes noticeable. It has already been shown[1] that 200 white lines spaced by 200 black lines in the height of the picture are 'clearly visible' at eight times the picture height by the average observer. All the data available from the present experiments indicate that, without receiver spot-wobble, the average viewer does not choose to come nearer than eight times the picture height to a 405-line picture.

The probable explanation of the rather revolutionary choice of standards suggested by the results of this investigation may lie in the observers' 'resistance' to the line structure in a single field when viewing a *picture*. When viewing a resolution chart the task is quite different and the irritating effect of the line structure is probably ignored. The introduction of receiver spot-wobble removes this irritation, and if sufficient bandwidth is present, as in our 405-line system, the viewer can move nearer to the picture and make use of this unbalanced excess of horizontal resolution on the basis of Baldwin's suggestion (Fig. 7) that a 2 : 1 out of balance either way in the ratio of horizontal to vertical resolution is tolerable. If this theory proves to be correct there is considerable room for improvement in the method of displaying interlaced television pictures, e.g. by the use of picture storage in the receiver. In the absence of a storage device, long-afterglow screens such as those used to demonstrate freedom from flicker[9] would appear to be worth investigation.

Some further viewer tests using an optical-interlace simulator are in hand in order to test the validity of the above theory. In this connection the frequency of the field scan may be more important than simple flicker considerations suggest. Thus the American 60 fields/sec system may appear much less 'liney' for the same line spacing as our 50 fields/sec system. For this reason the present results cannot now be applied by scaling to the American standards. This point is also under investigation with the interlace simulator.

The implications of the spot-wobble and synchronous spot-wobble experiments would appear to be as follows. If the optimum choice of standards is made in the first place, there would appear to be no direct advantage from the use of either. But with an excess of video-frequency bandwidth they will show a progressive advantage as demonstrated in these experiments. The present results may be considered as a preliminary indication of the possibilities, particularly of synchronous spot-wobble. Attention to spot-wobble waveform, frequency and amplitude in relation to the associated video-frequency bandwidth might result in still further improvements. The synchronous-spot-wobble system offers the opportunity to establish a television standard and then at some later date to inaugurate a higher-definition system on new channels using the same basic scanning standard, as suggested[10] in 1952. It is a short step to the addition of colour on a synchronous sub-carrier using the same 'burst' to lock the synchronous colour detector and the spot-wobble in the receiver. The use of such higher-definition pictures for cinema projection would be an obvious application. The standard signal would be sampled out in the manner demonstrated[2,7] for radiation over the normal broadcast television system.

With regard to colour television, the experiments show that the American N.T.S.C. proposal to put the chrominance information at the top of the video-frequency band is entirely justified on all existing 25-picture/sec standards. On the other hand, they show that it is not necessary to interleave the two signals. $1\frac{1}{2}$–2 Mc/s for 405 lines, 3–$3\frac{1}{2}$ Mc/s for 625 lines, and 6 Mc/s for 819 lines would appear to be ample luminance bandwidth. If these were adopted, chrominance information could be transmitted in the space vacated by the luminance, and in the same relation to sound and vision carrier as in the N.T.S.C. system, but without crosstalk, dot patterns and 'beaten-down'* noise. It would certainly be an instructive exercise to take the suggested optimum bandwidth for a system, say $1\frac{1}{2}$ Mc/s for a 405-line standard, or 3 Mc/s for 625 lines, and try to insert an N.T.S.C. type sub-carrier into it.

The effective video-frequency bandwidth appropriate to the viewers' requirements indicated by the present results gives greater importance to the colour information in the 405- and 625-line versions of the N.T.S.C. system so far demonstrated. For example, instead of considering the I and Q signals in the 405-line version of the N.T.S.C. system as being approximately one-third and one-eighth of a luminance channel of 3 Mc/s, they now have to be considered as occupying together about half the total available bandwidth, if the luminance is only effective to $1\frac{1}{2}$ Mc/s. However, this is quite consistent with the latest findings of Baldwin, who, in a recent publication,[11] indicates that a colour picture of equivalent sharpness to a monochrome picture can be produced by about 100% increase in the video-frequency bandwidth. In the light of the present experiments, therefore, European standards of television appear to be well suited to a separate chrominance allocation on this basis. With the trend to larger and brighter pictures the elimination of irritating edge effects on colour pictures would appear to be just as important as increasing the luminance definition.

An incidental advantage of the separate channelling of luminance and chrominance, by restriction of the luminance channel and without apparent loss of picture sharpness, would be the ability to run the television system locked to the mains supply rather than to the colour sub-carrier, as is necessary in the N.T.S.C. system.

The optical experiments show quite clearly that, for a given picture quality, people want to sit farther away from larger pictures in strict proportion to the picture dimensions. On the other hand, if quality is increased at the same time as size, and especially if it is increased in proportion to size (i.e. constant quality per unit area), the viewers remain at approximately the same viewing distance. This conclusion is borne out by both the optical and the television experiments.

The equating of television standards to cinema pictures has long been a pleasant pastime of television engineers. The last estimate in which the author was involved[1] gave 625 lines with receiver spot-wobble and 5 Mc/s video-frequency bandwidth as the minimum acceptable for a 4 : 3 picture when viewed at four times the picture height. The present experiments show this estimate to be at fault because such a picture would be viewed by choice at about $5\frac{1}{2}$ times the picture height. It may well be that a re-determination of the optimum viewing distance for a cinema picture would result in a choice nearer this value. The hitherto generally accepted figure of four times the picture height has been based on the position of the most expensive seats in the theatre, at the front of the circle. Figs. 8 and 11 show that at

* The ratio of the 'effective' number of vertical picture elements to the actual number of lines displayed in a television picture is known as the 'Kell factor', as a tribute to Ray Kell who did much of the original work on this subject.[8]

* Crosstalk noise produced by the sub-carrier of the N.T.S.C. system beating with the high-frequency noise in the luminance component of the signal.

$5\frac{1}{2}$ times the picture height 750 lines and 4 Mc/s would suffice. At four times the picture height 1 150 lines and 8·5 Mc/s would be required. If the chosen viewing distance were $4\frac{1}{2}$ times the picture height, for example, this reduces to 1 000 lines at 6·3 Mc/s. Probably the most useful conclusion that can be drawn from the present experiments is that a *maximum* of 8·5 Mc/s is required to equal a cinema picture (35 mm sound film; 4:3 aspect ratio). This is for 25 television pictures/sec and 2:1 interlace, of course.

The present trend in the cinema industry is towards a 2 : 1 aspect ratio coupled with larger screens giving the viewer more peripheral vision. In the Todd-AO and Cinerama systems this produces an effect now known as 'audience participation'. The general move towards larger viewing angles has immediately resulted in dissatisfaction at the sharpness of the pictures with those processes using normal 35 mm film-frame size. Considerable efforts are now being directed to producing finer-grain emulsions and better optical systems for cameras and projectors. The 'brute-force' method of projecting a larger area of film, however, is producing the most spectacular results. The area of the original negative projected per second has been increased $4\frac{1}{2}$ times for the Todd-AO system and seven times for Cinerama. This is exactly analogous to the method of the experiments shown in Figs. 3, 4 and 5, and, of course, to increasing the bandwidth and number of lines of a television picture. There is already some evidence that people are complaining of the picture 'lininess' on 21 in black-and-white receivers in this country, and that this size is about the limit for a 405-line system viewed in the larger-size living rooms. On the basis of the present experiments a change to 625 lines should allow this picture size to be increased to that of a 29 in tube, and for 819 lines to at least 34 in.

It may be encouraging to the cinema industry to note that, on the basis of the above reasoning, a Todd-AO picture would require 38 Mc/s maximum and Cinerama 60 Mc/s. The Teknirama negative, which is the largest known to the author for producing normal 35 mm prints for anamorphic projection, accommodates $2\frac{1}{2}$ times as much picture area as the normal 3:4 sound film, and is therefore equivalent to 21 Mc/s maximum.

The chosen value of V/H for the colour transparencies used in the direct-projection experiments was 4·1 (Fig. 2), corresponding to 8·1 Mc/s (Figs. 8 and 11). The area projected in these tests was 27·5 mm × 20·5 mm. The full area of such a colour transparency is 36 mm × 24 mm, corresponding, therefore, to 12·5 Mc/s. The chosen value of V/H for the 16 mm colour film was 5·5 (Fig. 5), corresponding to 750 lines, 3·9 Mc/s (or 625 lines, 5 Mc/s, with spot-wobble—see above).

A few incidental items arising from this investigation should be noted. The results explain why television recordings, and particularly the 'suppressed field' type in which only one field (half the total number of lines in the picture) is recorded, looks so good when projected normally on a viewing screen. Also the results explain completely why, in the course of the tests on the 405-line version of the N.T.S.C. system, it was frequently noted that there was no detriment to picture quality when the luminance channel was reduced to about 2 Mc/s either on the colour picture or on the monochrome compatible picture or even on normal B.B.C. reception.

The tri-colour tube when carefully aligned appears to be perfectly adequate for the 405-line N.T.S.C. picture so far as resolving power is concerned.

The following criticism may be levelled at the experimental method: The whole conception of the experiment is wrong, i.e. it does not tell us what we really want to know about picture quality, since observers may choose to sit at the same distance from two pictures and still prefer one much more than the other. This may be true as far as 'picture quality' is concerned, but when sharpness only is changed as in these experiments, and especially the number of lines and bandwidth of the television systems, there appears to be a very coherent relation with viewing distance.

It may be argued that the spot size was not adjusted either on the flying-spot scanner or on the monitors when changing from 405 to 625 lines. There seems to be no doubt that the spot was sufficiently small to deal adequately with 625 lines, in which case the picture might have looked excessively 'liney' when working on 405 lines. The fact that the results obtained on the two standards are so exactly related, as shown by Fig. 8, would refute this suggestion. At the viewing distances chosen it is almost certain that the apparent lininess would be governed by the total light flux in the lines and by their spacing, provided that the lines were fine enough.

It might be suggested that the effect of perspective on the viewers' judgment has been completely neglected in considering these results. From artistic considerations it has always been taught that viewers should sit at the same proportionate distance from a reproduction as the camera at the taking end. In other words, the viewer should accept the same angle of view as the camera did in taking the picture. It might be argued, therefore, that, in the optical experiments, when the picture was projected sharply focused, the viewers' reactions were entirely as would be expected on the grounds of accepting the same viewing angle as the taking lens. When the whole picture is projected at different sizes they move closer or farther away in proportion, as shown in Fig. 2. When progressively larger amounts of the picture are displayed to them they remain at approximately the same distance from the screen and accept the larger viewing angle, as shown in Fig. 4. This argument is completely refuted, however, by the optical experiments with the defocused projector. Either with the whole picture displayed, or with progressively larger parts of the picture, the viewers behaved as though they were reacting to picture sharpness entirely and not to perspective (Figs. 2 and 5). In the television experiments, the whole picture was used on every occasion and the sharpness was varied by the different television scanning standards. Again viewers behaved as if they were reacting to picture sharpness only (Figs. 6 and 8). However, it may well be that there is some second-order effect due to perspective, and this can be checked by repeating the experiments using a set of transparencies of the same scene, photographed with different focal-length lenses, and at proportionately different camera distances.

The results may be criticized on the grounds that they were obtained almost entirely with still pictures. The limited experience with moving pictures, however, did not indicate that any major correction would be necessary due to this factor. Three of the four colour transparencies had already been extensively used in America and in this country for assessing colour television standards.

It therefore appears that there is an excess of video-frequency bandwidth available—50% in the 405-line case—for all the European monochrome standards, and the question of how to make the best use of it requires consideration. The following ideas have been suggested.

(*a*) To investigate the possibility of increasing the apparent picture sharpness (reducing the chosen value of V/H) by receiver display storage, e.g. afterglow screens.
(*b*) To insert colour into the upper end of the video-frequency band but without interleaving.
(*c*) To increase the number of lines—405 to 625 and 625 to 819.
(*d*) To increase the picture ratio from 4:3 to about 2:1.
(*e*) Simply to allow the receiver manufacturers a cheap 'roll-off' in the video-frequency characteristic, without loss of picture sharpness.

It would appear desirable not to do more than one of these at a time.

In conclusion may it be said once again that, in establishing

television standards, and especially in the present instance colour-television standards, future developments in equipment must be envisaged just as they were for monochrome in 1935. In particular, potentialities in receiver development are very important, as it is here that most of the capital investment resides when a service is finally established. There seems no doubt that the ability to produce larger pictures in the home will materialize not too long after the establishment of a colour-television service in Europe. The experimental work described above indicates that, with present home-viewing distances, it will not be too uneconomical in bandwidth to provide the higher standard of definition required to go with them.

(6) ACKNOWLEDGMENTS

The author wishes to express his thanks to Sylvania-Thorn Colour Television Laboratories, Ltd., and to Marconi's Wireless Telegraph Co., Ltd., for permission to publish the paper; to his many colleagues at the Marconi Research Laboratory at Great Baddow for providing help, facilities and encouragement in the work; and to Mr. G. Downie for his patience and persistence in carrying out the observer tests.

(7) REFERENCES

(1) Jesty, L. C., and Phelp, N. R.: 'The Evaluation of Picture Quality with special reference to Television Systems', *Marconi Review*, 1951, **14**, pp. 113 and 156.

(2) Sarson, A. E., and Stock, P. B.: 'A New Approach to Telerecording', and Jesty, L. C.: 'A Review of Television Recording 1946–56', *British Kinematography*, 1957, **31**, pp. 119 and 87.

(3) Jesty, L. C.: 'Recent Developments in Colour Television', *Journal of the Television Society*, 1955, **7**, p. 488.

(4) Fredendall, G. L., and Kennedy, R. C.: 'Linear Phase Shift Video Filters', *R.C.A. Review*, 1950, **11**, p. 418.

(5) Baldwin, M. W.: 'The Subjective Sharpness of Simulated Television Images', *Proceedings of the Institute of Radio Engineers*, 1940, **28**, p. 458.

(6) Blumlein, A. D.: British Patent No. 503555. October, 1937.

(7) Marconi's Wireless Telegraph Co., Ltd., Jesty, L. C., and Sarson, A. E.: British Patent No. 758085. March, 1953.

(8) Jesty, L. C.: 'Horizontal versus Vertical Resolution', *Wireless World*, July, 1957. Kell, R. D., Bedford, A. V., and Trainer, M. A.: 'An Experimental Television System', *Proceedings of the Institute of Radio Engineers*, 1934, **22**, p. 1246.

(9) Haantjes and De Vrijer: 'Flicker in Television Pictures', *Wireless Engineer*, 1951, **28**, p. 40.

(10) Jesty, L. C.: 'Television as a Communication Problem', *Proceedings I.E.E.*, Paper No. 1336 R, April, 1952 (**99**, Part IIIA, p. 761).

(11) Baldwin, M. W., and Nielsen, G.: 'Subjective Sharpness of Simulated Color Television Pictures' (I.R.E. Transactions—Broadcast Transmission Systems PGBTS-4, March, 1956, p. 17).

(8) APPENDIX

Relevant Technical Data for the Experiments

(8.1) General

The size of the viewing room used in the optical-projection and flying-spot television experiments was 25 ft 8 in × 12 ft 3 in (approximately 8 m × 3·75 m). The screen was placed 4–5 ft (1·5 m) from the end wall remote from the door. Low ambient lighting was used. Observers were allowed time to accommodate before carrying out an observation. The pictures were displayed at a convenient height to be viewed seated.

The colour transparencies used in the optical projection and flying-spot television experiments were as follows: close-up of girl with scarf; materials and jewellery; autumn trees; Coronation decorations in Whitehall. The first three were from the N.T.S.C. set and the last was an original colour film.

The average age of the nine observers was 36 years. The standard deviation of their ages was ±10 years. Two were women. Four wore spectacles. Taking their individual V/H readings for each experiment, as a percentage of the mean in each experiment, and averaging for each observer over the whole series of tests (26 optical and 28 television) the following results were obtained. The three youngest observers (average age, 26 years) were 3·2% high on the average value of V/H; the next three (average age, 34 years) were 1·4% high; the three oldest (average age, 47 years) were 3·6% low.

(8.2) Optical Experiments

The projector had a 4 in f2·8 lens.

The area of colour transparency viewed in the sharp-focus condition was 20·5 × 27·5 mm².

The four viewing screens were coated matt white and were 12 in × 9 in, 17·3 in × 13 in, 24 in × 18 in and 34·6 in × 26 in.

Picture-highlight brightness was in the region of 20 ft-lamberts except where stated (¼ and 4 times brightness experiments). In the first series of experiments (Figs. 1 and 2) neutral density filters of 0·9, 0·6 and 0·3 were placed on the projection lens for the three smaller picture sizes, in order to equalize the brightness.

In the 'defocused' condition (Fig. 1) the projector was pulled back 2½, 5, 10 and 20 in, respectively, for the four picture sizes.

(8.3) Television Experiments

The low-pass filters used (see Table 1) were constructed according to Reference 4, with a transmission bandwidth of 6α, and a cut-off bandwidth of 1·5α.

The area of colour transparency scanned in the flying-spot scanners was 20·5 × 27·5 mm².

Tests on the 4½ in image-orthicon chain, including all the spot-wobble variations, were carried out with the observers standing viewing a rack-mounted monitor in the laboratory.

Details of the television displays are as follows:

405-line flying-spot colour scanner. 21 in tri-colour tube. Picture height, 14·5 in. Highlight brightness, about 10 ft-lamberts.

625-line/405-line flying-spot monochrome scanner, 21 in rectangular tube. Picture height, 13–14 in. Highlight brightness, about 20 ft-lamberts.

405-line 4½ in image-orthicon camera chain. 17 in rectangular tube. Picture height, 9·75 in. Highlight brightness, about 20 ft-lamberts.

Details of 83 home-viewing tests (Fig. 11) are as follows:

Picture height	Number of observers	V/H	σ
in			
6–8½	28	8·9	±28%
9–9½	29	9·2	±23%
10–13½	26	8·4	±27%
	Total 83	Av. 8·8	Av. ±26%

DISCUSSION BEFORE THE RADIO AND TELECOMMUNICATION SECTION, 19TH FEBRUARY, 1958

Mr. G. G. Gouriet: The author has described a number of carefully conceived experiments, and he has presented a convincing interpretation of the results. Nevertheless, I find it difficult to accept the somewhat startling conclusion that the optimum bandwidth for a 405-line television system is about 1·5 Mc/s. This and other similar conclusions rest on the validity of using the single parameter, viewing distance, as a means of rating picture quality.

The result of the first experiment shows that the preferred viewing distance increases linearly with the size of picture. The experiment was conducted in a room 26 ft long, and I wonder whether this did not give the viewers unrealistic latitude. We find that in practice, for one reason or another, the change of viewing distance with picture size is relatively small. This is borne out by results which we obtained during the recent colour-television compatibility trials, when 600 questionaries completed by monochrome viewers were analysed. Fig. A shows

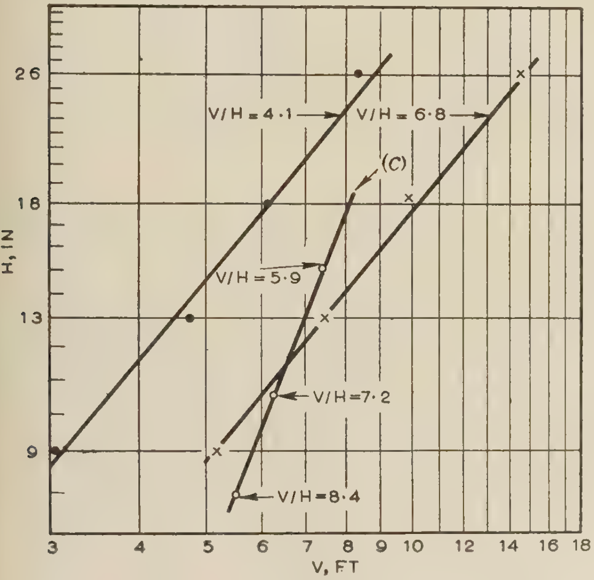


Fig. A

these results reproduced in Fig. 2 of the paper as curve (*c*), and it will be seen that the slope of this curve is substantially greater than that of the author's curves indicating that, in practice, the range of viewing distance is fairly restricted. It is strange that the author analysed 83 questionaries from the same series of tests and obtained an average of 8·8 for the ratio V/H; this is notably higher than the average ratio obtained from the 600 questionaries.

I should like to comment briefly on the filters which the author used to restrict the bandwidth of video-frequency signals. In television, where a phase-equalized filter is normally used for this purpose, it is reasonable to regard the effective bandwidth as being that of an ideal filter having the same area under the response curve. Such an ideal filter will give the same maximum slope in its response to unit step and, therefore, a similar degree of picture sharpness. On this basis, the filters used for the experiments each have an effective bandwidth 10% higher than the figure quoted in the paper.

The author has suggested that, in practice, more efficient filters could be used, but here I disagree. In using a bandwidth of Mc/s for a 405-line picture it is permissible to use a high rate of cut-off, since any 'ringing' produced will have a relatively fine structure of low visibility. On the other hand, if the bandwidth is abruptly restricted to 1·5 Mc/s, 'ringing' will produce a pattern twice as coarse as the line structure in a single field; I believe that this is more objectionable than is the visibility of scanning lines.

In order to avoid 'ringing', the cut-off must be gradual and for the same *effective* bandwidth this inevitably means using a greater total bandwidth. This argument is equally applicable to 625 lines when the bandwidth is limited to 3 Mc/s. I should be interested to know whether the filters used by the author did in fact cause 'ringing', and if so, whether the viewers commented on the fact and objected to it.

I am very interested in the results which the author has obtained using spot-wobble techniques. It should be mentioned that photographs can be misleading, in that they do not show line break-up and stroboscopic effects. Our experience with interlaced sampling, which is not unrelated, has been that, whilst photographs show a marked advantage, when directly viewed the picture is disappointing.

Mr. I. J. P. James: Mr. Gouriet's remarks concerning a bandwidth of 1·5 Mc/s remind me of tests we did in the early days of the development of the C.P.S. Emitron pick-up tube. During transmissions from Wimbledon we were testing various networks in a camera channel. The observers at the receiving end preferred the pictures when a single constant-K section low-pass filter with a cut-off frequency of 1·5 Mc/s was inserted in the channel. The observers preferred a narrow-bandwidth picture to one with a wider band accompanied by high-frequency noise. The author does not mention the effect of noise, but this should be considered in assessing bandwidths and the quality of pictures.

With regard to the tests shown in Fig. 6, in comparing 625 lines with 405 lines one wants to be quite sure that one is testing the viewer's reactions and not the cathode-ray-tube characteristics. It appears from the photographs in the paper that the line structure was rather finer than it should have been. In other words, the 625-line pictures have the gaps filled more than the 405-line pictures, and this tends to accentuate the difference between the two systems.

If we consider a sequential television system in which we have ideal square-section spots, so that the lines are jutting against each other, we shall not see any line structure. Therefore, it does not matter how many lines we have; we cannot see them. If we interlace them, and the eye moves, the raster tends to split up even with ideal spots; thus interlacing does introduce trouble, which is presumably the effect, together with inter-line flicker, that is mainly shown by the curve. It is very important in tests to make sure that the rasters are interlacing correctly; and the author has obviously taken this into account.

Some experiments similar to those described in the paper are reported by F. D. Thompson.* Fifty viewers were used to determine the distance from a 24 in receiver at which the line structure became barely resolvable. The test was conducted on a standard 525-line raster with no video-frequency modulation and a brightness of 20 ft-lamberts. The viewers backed away from the receivers until the lines just blended together. The distances were recorded for the conventional raster and for a 13·25 Mc/s wobbled raster.

The average viewing distance for the raster was 10·6 ft (vertical viewing angle of 7·6°), giving $V/H = 7{\cdot}5$. The average distance for the wobbled raster, 6·1 ft (13·1°), gives $V/H = 4{\cdot}3$. In additional experiments, some of the viewers were seated in chairs with castors and were asked to move about and pick the location from which they preferred to view the receiver. They went to more or less the same positions as they had chosen before.

If the values for V/H of 7·5 and 4·3 are plotted in Fig. 11 for a bandwidth of 3·4 Mc/s (4·2 Mc/s less 20% because of the

* THOMPSON, F. D.: *Journal of the Society of Motion Picture and Television Engineers*, 1957, 66, p. 602.

difference in field rate), there appears to be fair agreement for the conventional raster, but a considerable discrepancy for the wobbled raster.

Mr. T. Kilvington: The author's method of assessment is an extremely interesting one which appears to be capable of taking into account all the factors that affect the subjective quality of a picture. Perhaps for the first time it provides a method of taking the line structure into account. In the past, line structure has been taken for granted, and what has been assessed has been the horizontal definition. But I notice that in the tests described a still picture was used and great care was taken with the interlacing. In viewing moving pictures, when there is some vertical movement in the picture or some vertical movement of the camera, the interlace frequently seems to disappear completely and a picture with half the number of lines is seen. It would be interesting to know whether any experiments with moving pictures were made to assess this effect and, in particular, whether any comparison has been made between an interlaced system in which this effect occurs and a sequential system in which it does not.

The method would also appear to be capable of taking into account the effect of noise, for it is an undoubted fact that in the presence of noise a picture will look more acceptable from a greater distance. Has this possibility been investigated?

I was surprised to find that the author could gain any support for his theories from the results of viewing tests in people's homes. It seems to me that viewing distance at home is determined not so much by the cathode-ray-tube size as by other factors such as the size of the room, the number of people viewing simultaneously and the disposition of the furniture within the room. Speaking from experience at home over the last ten years or so, although we have graduated in stages from a 6 in tube to a 21 in tube, I am quite sure that our average viewing distance has not increased in the same ratio of 3·5 : 1.

With regard to spot-wobble, or rather an alternative to it, some years ago we made a monitor in which particular care was taken of focus and interlace so that the line structure was clearly visible. We found it possible, with the aid of two small pieces of magnetic material inserted in the focusing-magnet system, to distort the focusing field in such a way that the spot became elongated in the vertical direction. This broadened the lines, reducing their visibility and giving an overall improvement in picture quality. It seems that some static method of line broadening such as this may be a simpler and cheaper method of reducing line visibility than the better-known spot-wobble technique.

Mr. B. C. Fleming Williams: It is about a quarter of a century since the present television standards were proposed. They have stood the test of time remarkably well, and we all owe a debt of gratitude to the people who devised them for doing such a fine job. With new techniques and larger display tubes a reconsideration of the standards is timely. The work done by the author provides another approach to the problem of making subjective measurements in an objective manner, and his paper is in this, and other ways, a valuable contribution.

When further work is done along these lines, I think care is necessary to ensure that the people used are not previously conditioned by habitually viewing a certain size of tube. A number of children should also be included in the tests, as, in my experience, these tend to sit closer to the display than their elders, and, if this were proved to be general, an age factor would have to be considered in choosing future standards.

Colour television will have to come some time in the near future, and this forces us to reconsider television standards in any case.

Mr. S. N. Watson: I have read the paper many times and I find that within itself the evidence presented is consistent and logical. But three important factors have either been omitted or need modifying.

First, the author has shown his viewers an insufficient range of types of picture. I can present evidence to show that, in a restricted bandwidth, one can, with certain types of picture, actually lose information which no moving about in the viewing distance can replace. Secondly, the author's estimate of the equivalent rectangular bandwidth of his filter is too low. The third, and by far the most important, point is how much of the nominal bandwidth of a television system reaches the viewer in his house. My own opinion is that the average rectangular bandwidth which the British viewer gets is around 2 Mc/s. This is no criticism of the designers and makers of the British television receiver. It is just a combination of practical factors which produces this kind of answer.

My own guess is that if you take these three factors together you come to something like twice the bandwidth estimated as necessary by the author. That brings us back more or less to the television standards at present in use in Europe.

There is also a fourth factor which is related to the three mentioned already. This is the effect of distortions on the quality of pictures. The author has carried out a laboratory experiment, ignoring the difficulties of getting the laboratory picture to the home viewer; such things as noise, distortion due to vestigial reception, quadrature distortion, echoes on the picture, etc., have been omitted. An attempt to use a low bandwidth for a high-speed scanning system would make such distortions very visible.

To add to the statistics, I have the distances adopted by viewers in our recent colour tests. There were 300 observers using 21 in receivers, the vast majority in homes. 9% adopted a value for V/H of 3·5, 65% of 5·5, 22% of 7·5, 3% of 9·5 and only 1% greater than 9·5. The average of all taken together was 6.

Mr. R. C. Winton: The paper is interesting because it deals mathematically with something which has hitherto been mainly empirical. For a long time we have worked to a figure of ten times the height for the ideal viewing distance, and it is surprising that this is so near the figures quoted.

Taking a total of 30 million viewers, the sampling rate on which the observations are based is some 3 million to 1, which seems remarkably low. How were the observers chosen? Was anything known of the state of their eyesight or whether they were habitual viewers or cinemagoers?

I doubt whether there is a public requirement for larger pictures in the home. The 17 in set still constitutes 70% of the market, although the 21 in tube has been available for three or four years. The latter is less popular chiefly because of the size of the cabinet and the high price, and I am very doubtful whether it will ever be more popular than the 17 in tube.

The practical difficulty of spot-wobble is to prevent the wobble frequency interfering with the picture. The coil must radiate to do its job, and it is difficult to contain this radiation in the right place. What suggestions has the author for incorporating spot-wobble into a marketable set which will not be too expensive?

Dr. D. A. Maurice: The author has used his results as an attack on bandwidth: I should like to use them as an attack on interlacing. I think it has been greatly overrated; probably it was more effective in the early days when display screens were less bright than now.

I want to refer to a paper* by Dr. Gilbert in which correlation between visibility of picture detail and visibility of grating patterns was established. Using black-and-white test charts in an optical experiment, Dr. Gilbert found that, for a grating pattern having a half-wavelength equal in dimension to that of the gap in a Landolt ring, the threshold of visibility occurred at a distance equal to four-thirds that for the threshold of visibility

* GILBERT, M.: 'Definition of Visual Acuity', *British Journal of Ophthalmology* 1953, 37, p. 661.

of the broken-ring detail, which other experimenters have found to be in accordance with detail in photographic reproductions. Dr. Gilbert then used a 405-line 6 Mc/s sequential television system and she found that, this time, the threshold viewing distance for the raster was approximately three-quarters of that for the picture detail. This is because the picture detail in such a television system, reckoned in terms of either horizontal or vertical resolution, has a dimension approximately equal to a wavelength of the raster pattern rather than a half-wavelength, as was assumed in the optical case. A 405-line 3 Mc/s interlaced system was next used, and in this case the threshold viewing distance for the raster was approximately 3/2 times that obtained for the picture detail. This drastic change in the ratios of the threshold distances is due to the greatly increased visibility of the interlaced raster owing to the fact that the eye is susceptible to flicker at 25 c/s when normal present-day viewing screens are used.

Dr. Gilbert's tests have shown beyond doubt that, although it is possible to enjoy to the full all the detail in a sequential television picture without seeing the raster, it is not possible to do so with an interlaced raster.

It would seem from the author's work that, when given the opportunity, viewers would sit at such a distance from the receiver that they would be unable to see the (interlaced) raster, and at such a distance, therefore, the horizontal resolution supplied by all present-day television systems would be unnecessarily great.

Interlacing should be regarded as a low-cost fractional improvement which may be added to an otherwise adequate sequential television system.

Mr. P. P. Eckersley: There appears to be a suggestion in the paper that it would be possible to reduce the bandwidth in which television programmes are contained without serious loss to the viewer. The reduction of bandwidth must mean that less information would be conveyed, and while we may make the better of a bad picture by moving away from it, it would seem preferable to seek to improve the picture and invite its nearer view.

Mr. E. Ribchester: I have carried out some measurements along the same lines as the author.

Fig. B was produced using three 405-line standard receivers,

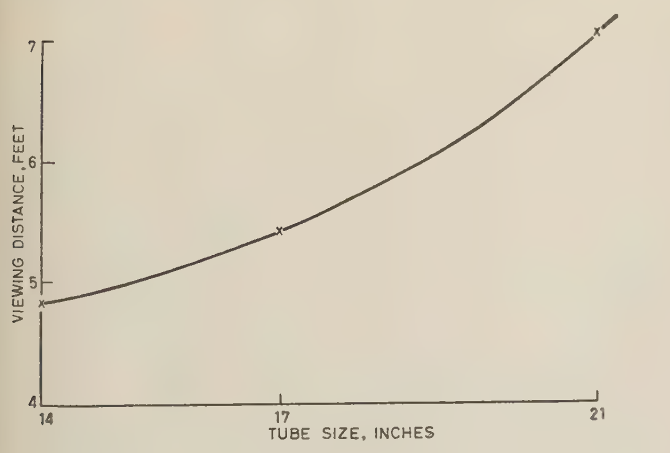


Fig. B

having 14, 17 and 21 in tubes. However, there was a slight difference which will be considered later. If you work out the ratio of V/H it is found to vary from about 5·7 to about 6·8. These figures are slightly lower than those shown by the author. There was one additional feature which was of interest, i.e. one viewer insisted on sitting about 20° off the axis.

We now come to the significant point about the measurements. During the experiments none of the receivers was switched on. This, I think, proves the point that you can produce consistent results, although they do not mean very much except perhaps that viewers have definitely conditioned themselves to 405-line pictures, probably of very doubtful quality. On examining the results in detail, I found that the people who were known to have the worst receivers at home sat the furthest away from the sets in the laboratory.

Mr. W. N. Sproson: I should like to report briefly about two experiments. In the first, we attempted an appraisal of 405-line monochrome television. The receiver was a 21 in high-grade commercial one, 96 observers were used and the picture material was in the form of four outside broadcasts of sporting events. The observers were seated at four, six and eight times the picture height. The results show that the observers saw the scanning lines quite clearly, and furthermore, there is good negative correlation between line visibility and viewing distance. A slight extrapolation of these results gives the figure of nine times the picture height as the distance at which the lines are just perceptible. The figure of nine agrees well with the point of maximum rate of change of slope in Fig. 6 of the paper. I suggest that this is the parameter which the experiments yield, namely the distance at which the lines disappear.

Correlation analysis applied to the 96 sets of results showed that there was no correlation between line visibility and overall assessment of picture quality. It would appear that, in this experiment, the line visibility was ignored by the observers and their picture appraisal was unaffected by it.

The second experiment relates to the region between 1·5 and 3 Mc/s, which the author thinks is of doubtful value. The purpose of this experiment was to determine the equivalent rectangular bandwidth of a picture which is just perceptibly degraded as compared with a fully-resolved 3 Mc/s picture. The observers were seated at four times the picture height and the result was an equivalent rectangular bandwidth of 2·67 Mc/s. Thus, on those occasions when people do sit fairly close to a television screen, a bandwidth considerably in excess of 1·5 Mc/s is required.

Mr. D. C. Birkinshaw: I have two comments on the paper. First, I join those who are drawing attention to the importance of the line structure in this investigation. The line structure must be categorized as a disability of most television pictures which we see to-day. One may not unfairly describe the current British television picture by saying that the transmitting authority radiates a 405-line interlaced picture which is displayed as an 810-line interlaced picture consisting of the 405 lines of wanted information as transmitted interlaced with 405 lines of noisy black level. Thus the pictures possess a very obvious line structure which must be regarded as interference and eliminated as we eliminate other forms of interference.

Any observer employed in the author's experiment must surely have been influenced in selecting where he would sit by the undesirable presence of the line structure. I would therefore have thought that before the author conducted any experiments calculated to inform himself about the relation between viewing distance and picture quality, he would have cleared away the line structure. I agree that he has not ignored the matter in his special television experiments (indeed he mentions the resistance of observers to line structure in Section 5), but he does not consider it in his main experiments which preceded those in the 'special' category.

My second point concerns the Kell factor. This was developed as a coefficient in the formula for calculating bandwidth and intended to take account of the statistical distribution of fine detail in the ordinary run of pictures. The natural value of the Kell factor is 1·0, and the lower you make this factor, the more you are assuming that there will be an appreciable proportion of the pictures having detail so distributed as not to require

the full bandwidth. The lowest figure that I have previously seen suggested for the Kell factor is in the region of 0·65, and I should have thought that the value proposed by the author, i.e. 0·42, was distinctly low and that the statistical detail distribution in normal television pictures could not support a Kell factor as low as this.

Mr. P. J. Hewitt (*communicated*): While associated with the transmission of television signals, I naturally did much programme monitoring, and, in general, I agree that the distances measured by the author are typical for trained personnel.

Among the general viewing public, however, I think that the points raised by speakers have a far greater effect on the location of viewers than technical considerations, which are seldom appreciated by the viewer anyway.

Considering the pictures I see on my friends' receivers, with their incorrect aspect ratio, poor focus, sound on vision and 'soot and whitewash' effect, all of which are readily tolerated and eventually accepted as normal, I am convinced that a reduction in bandwidth would be noticed by only a few. It is the novelty and not the technical quality of a programme that attracts viewers; this is borne out by the success of the Eurovision transmissions, many of which had appalling definition. The average programme surely does not make full use of the allotted spectrum, and the only programmes that do must be the really high-definition films that are shown from time to time.

Mr. M. W. Baldwin (*United States: communicated*): In the optical experiments, consistent and reproducible values of the viewing ratio V/H were obtained when the colour transparencies were sharply focused on the screens. It would be interesting to know what determined the viewing ratio in this 'sharp-image' case, either from statements of the observers or from the author's knowledge of the apparatus. Is it possible that the 'sharp' image was sufficiently 'unsharp' to be controlling? I think that it probably was not, and I would like to know about more subtle factors, like perspective and home-viewing habits. My real concern is for assurance that viewing ratio is a proper criterion of sharpness.

I notice that the viewing ratio increases by about 50% when the projector magnification is doubled, for constant screen height (26 in) and constant brightness (one-quarter normal). This same increase is also evident for the 13 in screen, although in this case the brightness drops from four to one times normal. Why should the viewing ratio depend upon magnification?

In the experiments shown in Fig. 6 of the paper, there is clear indication that the author's observers and my observers (Fig. 5 of Reference 5) were not responding to the same aspect of picture quality. The author's observers increased their viewing ratio by about 30% when the number of lines changed from 625 to 405, for bandwidths greater than 3 Mc/s. My observers reported a change in subjective sharpness of less than 0·5 liminal unit under corresponding conditions (change from 8/5 to 5/8 for the upper two curves of my Fig. 5). A change so close to threshold could hardly be responsible for a 30% change in viewing ratio. Therefore, I am afraid I cannot agree that my Fig. 5 provides any confirmation for the view that there was a 'marked maximum' when the ratio of vertical to horizontal resolution was varied around the equal-resolution condition.

This leads me to wonder whether the juxtaposition of optical and television data, as in Fig. 6, is a reliable guide to numerical conclusions about scanning standards. Indeed, the same conclusions are drawn (from Fig. 10 instead of Fig. 8) without using any optical data at all.

In my opinion, the only evidence that viewing ratio is a proper criterion of sharpness comes from the optical experiments, and I wonder whether it is good evidence. The viewing ratio changes almost as much in response to a doubling of the magnification as it does in response to the pulling back of the projector from sharp focus. (Viewing ratio increased 50% for doubling the magnification, and increased only 70% for pulling back to a quality 'somewhat better than a good 405-line television picture'). I do not think that doubling the magnification should reduce the sharpness to that extent. This explains my need for assurance that viewing ratio measures what you want it to, without undue influence from magnification, field size, perspective, experience, or any other factor not directly related to scanning standards.

THE AUTHOR'S REPLY TO THE ABOVE DISCUSSION

Mr. L. C. Jesty (*in reply*): Messrs. Gouriet and Kilvington suggest that home-viewing conditions cannot reproduce the results obtained in the present experiments. Nevertheless their effect can be seen. Viewers frequently reject 21 in receivers in favour of 17 in ones because the picture proves 'too liney'. A change to 625 lines would permit 27 in receivers to be exchanged for 17 in ones without discomfort (see Fig. 2 of the short review* of the paper). This is fully supported by experience with u.h.f. 625-line tests currently radiated by the B.B.C.

As expected, the greatest discrepancy between Fig. A and the present data is for the largest picture size. The largest recorded in the 83 questionaries was 13½ in, and there were very few of these. Mr. Gouriet gives data for a height of 15 in. This surely represents a very small number of observations and is not likely to be reliable.

Messrs. Gouriet and Watson discuss filter characteristics. As pointed out at the end of Section 4 this is a matter for further investigation. If the values used in the paper are increased by 25%, which should eliminate ringing, the optimum 1·1 Mc/s bandwidth for 405 lines given by Fig. 6 still does not reach 1·5 Mc/s.

Messrs. Watson and Kilvington comment on the limited range of pictures used. When the data were broken down 'picture wise' there did not appear to be any significant difference due to picture content.

Dr. Maurice, supported by Mr. James, makes some very potent comments about the deficiencies of interlacing. I heartily agree. As pointed out, attention to decay characteristics of receiver screens and future receiver storage displays may yield great benefit to picture quality on existing standards. Mr. Birkinshaw's comments relate to this point. One cannot eliminate the line structure in considering television pictures. The 'natural' value for the Kell factor is more like 0·7 than 1·0 owing to the quantizing effect of the lines. If these are interlaced, apparently it is still further reduced.

A number of speakers have commented on the spot-wobble experiments. The paper referred to by Dr. Thompson supports the present findings, but the effect of the American 60 fields/sec system on persistence of vision may be more important than has hitherto been realized. The request for a simple non-radiating spot-wobble is met by Dr. Thompson's split-anode tube. Cylindrical defocusing has not proved very effective, probably because spot-wobble can be adjusted to give each line the appearance of two [see Fig. 30(*c*) of Reference 1], thus eliminating line crawling.

I am glad to have Mr. James's support for a 1·5 Mc/s picture. His suggestion that line structure might be too visible in the

* *Journal I.E.E.*, 1958, 4, p. 189.

405-line experiments is met by the spot-wobble tests, which did not approach 625-line quality even at 3 Mc/s bandwidth.

Mr. Watson discusses the desirability of allowing for some loss in transmission. It is standard motion-picture practice to produce a 'perfect' negative and allow the maximum tolerances in subsequent processing and projection. This is equivalent to the 'cheap roll-off characteristic' referred to in Section 5(*e*). Standards should be chosen so that expensive receivers can realize the full performance. Supply and demand will settle the lower limit.

Mr. Winton queries the reliability of relatively small numbers of observers to represent 30 million. It is a sad reflection on our (lack of) individuality that they have already proved to be so in numerous similar instances. Of course, one has to avoid choosing a team with fervently dedicated views.

Mr. Fleming-Williams points out that the advent of larger pictures opens the way to higher standards of definition. Already the 110° tube is eliminating the receiver cabinet as a controlling factor. Mr. Eckersley wants better pictures in preference to economical channels. I am delighted. With his penetrating eye he has unmasked my objective.

I am greatly intrigued by Mr. Ribchester's experiment with blank screens, and delighted to find confirmation that viewing distance increases with size, when information remains constant.

Mr. Sprosen's experiments used a fixed viewing distance. Our own showed repeatedly the zest with which observers pulled their chairs closer to a well-defined picture. It is now difficult to see the value of maintaining a fixed distance, which inevitably results in unrealistic saturating liminal characteristics. Close viewing of 405-line pictures obviously needs more than $1 \cdot 5$ Mc/s bandwidth, but it would appear that the average viewer does not want to sit close to such a picture even with unlimited bandwidth.

Mr. Hewitt stresses the overriding importance of programme material, and once the television receiver is installed this is undoubtedly true. Technical standards have their place, however, and the average man has no difficulty in choosing a receiver which gives a superior performance.

Mr. Baldwin raises two pertinent queries—what are the effects of perspective and habit? The height of picture used on the colour transparencies was $20 \cdot 5$ mm. The 'Whitehall' picture was taken with a 50 mm lens, and it is highly probable that the three N.T.S.C. pictures were similarly photographed. The value of V/H for correct perspective would therefore be $50/20 \cdot 5 = 2 \cdot 5$. Values obtained in the experiments range from $4 \cdot 1$ (Fig. 2) to more than 18 (Fig. 6). It seems unlikely, therefore, that perspective played any serious part. Cinema and television viewers are, in fact, quite acclimatized to wide changes. I have never observed any tendency for them to push their chairs back when suddenly presented with a very foreshortened view of a cricket pitch. With regard to habit, 405-line pictures are apparently viewed at eight to nine times the picture height in the home. The fact that the experimental data are so self-consistent and that a good 625-line picture or sharp optical projection gave values very much less than this shows that, if anything, observers' habits are governed mostly by the information content in the picture. At the completion of the experiments a questionary was circulated asking the observers how they arrived at their chosen viewing distance. Up to this time no suggestion had been made to them. They had only been invited to try going nearer and further away. Their replies contain such statements as 'reconciling viewing angle with definition', 'balancing viewing angle and sharpness', 'tended to stand further back when the picture was blurred or the lines showed more clearly'.

Owing to the necessity for compressing data for publication the derivation of Figs. 6, 8 and 10 may have been over-simplified. Fig. C shows this in more detail. The curves are the best fit to the experimental points and the flying-spot and image-orthicon

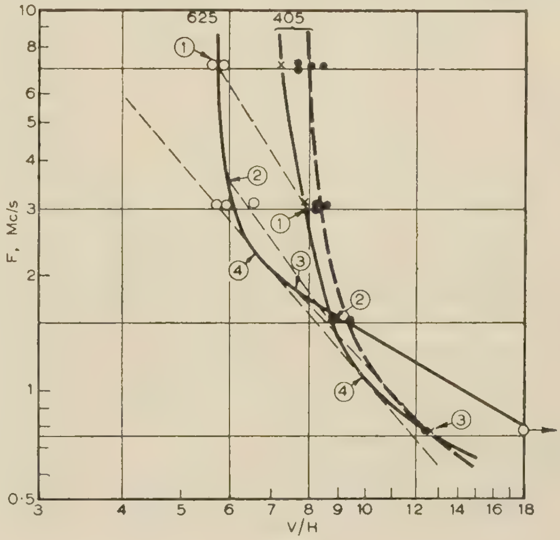


Fig C

---- $4\frac{1}{2}$ in image orthicon chain.
—— Flying-spot scanner.

data for 405 lines have been separated. Four dotted lines are shown joining 'corresponding' points on the 625- and 405-line flying-spot curves. The bandwidth ratio of each pair was been deliberately chosen to be $(625/405)^2$, so that, whatever the ratio of horizontal to vertical resolution, it remains the same for each pair of points. That is to say, the definition has been scaled up exactly in proportion horizontally and vertically in going from 405 to 625 lines along each dotted line. This is analogous to the second series of optical experiments (Fig. 5). It will be seen that the slope of the four dotted lines is practically the same, so that, whatever the absolute ratio of horizontal to vertical resolution, the observers always move nearer in the same ratio. The average 'slope' is $-2 \cdot 21$. The optical experiments (Fig. 5) gave $-2 \cdot 46$. By carrying out the procedure used to derive Figs. 8 and 10, and varying the slope of the tangent line deliberately, an estimate was made of its most probable slope on the basis that this would give the maximum number of points falling inside the chosen statistical limits. This was $-2 \cdot 2$. If this lower value is chosen for the television experiments, the optimum bandwidth for higher-definition systems becomes even lower than indicated.

The suggestion that Mr. Baldwin's observers and our own were not responding to the same aspect of picture quality is the subject for a paper in itself. Both experiments showed that, as the ratio of vertical to horizontal resolution is changed, viewer reaction goes through a maximum, but more sharply defined in the present experiments. This greater sensitivity of the observers is probably due to their being free to move rather than appraising the pictures at a fixed distance. As picture quality is progressively increased, they come steadily closer without apparent limit, whereas if they are viewing at a fixed distance, their grading of the picture rapidly reaches saturation. It seems obvious that variable viewing distance should give greater sensitivity and a smaller limen. Some experiments recently initiated indicate that such a limen is less than half that obtained from fixed viewing experiments.

621.373.44 : 621–397.6

The Institution of Electrical Engineers
Paper No. 2687 R
Sept. 1958

A PULSE-AND-BAR WAVEFORM GENERATOR FOR TESTING TELEVISION LINKS

By I. F. MACDIARMID, Associate Member, and B. PHILLIPS, B.Sc.(Eng.).

(*The paper was first received* 29*th January, and in revised form* 8*th May,* 1958.)

SUMMARY

Among the test signals used for measuring the transmission performance of television links are 'sine-squared' and 'smoothed-bar' pulses. The paper gives details of the design of a test-signal generator which, at the line-frequency repetition rate, produces a composite waveform consisting of a sine-squared pulse, a smoothed-bar pulse and a normal line-synchronizing pulse. The sine-squared pulse may have a half-amplitude duration of either T or $2T$, as selected by a switch, where $1/2T$ is the nominal upper frequency of interest in the television system (e.g. 3 Mc/s in the British 405-line system). The bar pulse has a duration of approximately one-half of the line period and has transitions of 'integrated sine-squared' shape.

To facilitate precise measurements the generator has been designed with particular emphasis on stability of waveform shape and freedom from jitter.

Consideration is also given to analogous test signals for colour-television systems.

(1) INTRODUCTION

Setting-up and maintenance operations on point-to-point television links are simplified and put on a more rational basis if the video transmission performance limits are specified in terms of waveform responses instead of steady-state attenuation/frequency and phase/frequency characteristics. A method of placing limits on the responses of a link to a standardized set of test waveforms has been described by Lewis[1] and this method has been successfully adopted by the Post Office with a consequent reduction in both setting-up and maintenance costs. In this method restrictions are placed on a number of different features of the waveform response and the limits are expressed in terms of a rating factor, K, whose numerical value is chosen to suit the degree of stringency appropriate to any class of link. The test signals required include a sine-squared pulse of specified duration and a 'smoothed' half-line bar. After indicating briefly some of the properties of such signals, the paper describes a generator for producing them in a stable and reproducible manner.

The properties and advantages of the sine-squared pulse as a test signal for television links have been adequately described previously,[2–4] and it is sufficient to say here that the sine-squared pulse of half-amplitude duration T, where T is the reciprocal of twice the nominal upper cut-off frequency of the link ($0 \cdot 17$ microsec for a 3 Mc/s link), provides a convenient signal for measuring distortions at the upper end of the frequency band. For routine tests, a pulse of half-amplitude duration $2T$ ($0 \cdot 33$ microsec for a 3 Mc/s link) is also desirable, since a pulse of this duration should be transmitted without change of shape. This property simplifies the use of masks for the delineation of distortion limits.[1]

A step-function type of signal is also required for the measurement of distortions in the lower part of the frequency band, and it is convenient to use for this purpose a rectangular pulse or 'bar' waveform whose duration is approximately equal to one-half of the effective line period. For most purposes it is desirable that the shape and rise time of the bar transitions should be controlled.[5, 6] A suitable transition has the integrated sine-squared shape defined by

$$V(t) = (1/\tau)\int_0^t \sin^2 (\pi u/2\tau)du$$

where $0 \leqslant t \leqslant 2\tau$, τ is the half-amplitude duration of the corresponding sine-squared pulse, and u is a variable of integration. The rise time of the integrated sine-squared transition between the 10% and 90% points is $0 \cdot 96\tau$. A bar pulse with transitions of this shape can be generated by passing a rectangular pulse of suitable duration through a sine-squared pulse-shaping network of the type previously described for sine-squared pulse generation,[2] the response of such a network to a step function being the integral of its impulse response. To ensure that the transitions are controlled by the network, the rise time of the input pulse should be short, preferably not greater than $\tau/5$. The calculated impulse and step responses of the network used are shown in Fig. 1.

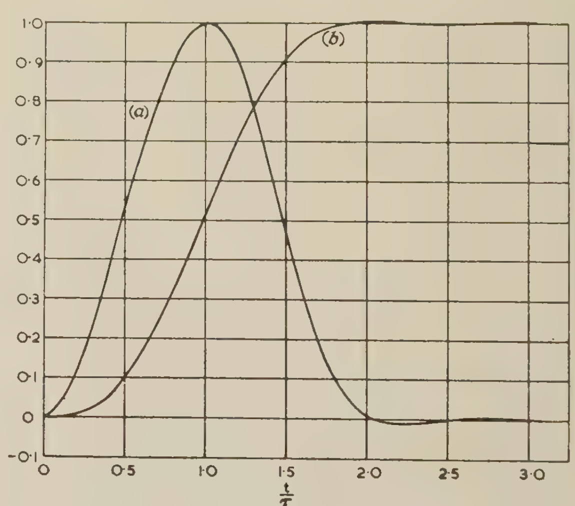


Fig. 1.—Calculated response of sine-squared pulse-shaping network.
(*a*) To ideal impulse.
(*b*) To ideal step.

It is obviously economical to use the same network for shaping the sine-squared pulse and the bar edges—a fact which suggests the generation of a combined waveform. In addition to being economical, the combined pulse-and-bar waveform has the important practical advantage of facilitating the measurement of changes in the relative amplitudes of pulse and bar caused by transmission over an imperfect link. This form of distortion, which is one of the features measured in the determination of the rating factor of a link,[1] corresponds to an error in the brightness with which fine detail is reproduced in a television picture as compared with the brightness of large areas of uniform tone. In addition to its use as a measure of link distortion, the change in bar/pulse amplitude ratio has been found to be a

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Mr. Macdiarmid and Mr. Phillips are at the Post Office Research Station.

particularly useful and sensitive criterion in the alignment of certain types of video amplifier and in similar applications.

On certain types of television link[7–9] line-synchronizing pulses are required for supervisory or control purposes and consequently should be included with the test waveform. Provided that they can be generated so that they are free from jitter relative to the other parts of the waveform, they also provide the most satisfactory method of starting the oscillograph sweep in sufficient time to permit the display of all the distortion which may occur before the main lobe of the sine-squared pulse or either transition of the bar.

The paper describes a pulse-and-bar waveform generator for use with the 405-line 3 Mc/s television system, which has been designed to provide a waveform consisting of three interlaced pulse trains, each with the same repetition frequency (10 kc/s) and accurately locked in time to avoid jitter. The pulse trains are (*a*) 10 microsec synchronizing pulses, (*b*) sine-squared pulses of half-amplitude duration T (0·17 microsec) or $2T$ (0·33 microsec), and (*c*) bar pulses of 40 microsec duration, with edges of integrated sine-squared shape controlled by the networks used to shape the sine-squared pulses. The generator delivers a waveform of 1 volt d.a.p. into a 75-ohm load, with the usual 70 : 30 ratio between picture and synchronizing signals. The output impedance of the generator is also 75 ohms, and variations of load impedance do not affect the shape of the generated waveform.

(2) OUTLINE OF CIRCUIT ARRANGEMENT

One requirement for the generator which plays a dominant part in determining the form of circuit arrangement used is that the synchronizing pulses should be suitable for triggering an oscillograph time-base which may be adjusted to examine any part of the waveform. This imposes a very severe restriction on the amount of jitter which can be permitted in the interval between the synchronizing pulse, the sine-squared pulse and both edges of the bar pulse. The maximum sweep speed required for use with this waveform is about 10 cm/microsec on a normal-sized cathode-ray tube. If the effective line width of the trace is about 0·2 mm, the maximum jitter of the signal which will not seriously increase the line width is about 1 millimicrosec. This time stability must be maintained for delays up to about 80 microsec from the synchronizing pulse and so the short-term time-stability required is about 1 part in 10^5. This timing accuracy is required at five points in the complete waveform which can all be separated by integral multiples of 10 microsec. A 100 kc/s oscillator of good phase stability is therefore used to generate a train of steep-fronted pulses at 10 microsec intervals which determine the timed points of the output waveform. The continuous train of pulses at 100 kc/s is separated into five trains at 10 kc/s repetition frequency by a time-selection process, and these trains are used to control the timing of the output waveform. The method can now be described in greater detail with the aid of the block diagram in Fig. 2 and the timing chart in Fig. 3.

The output of a 100 kc/s *LC* oscillator triggers a narrow-pulse blocking oscillator which has two outputs. One of these outputs drives (via a diode gate which suppresses every ninth pulse) a blocking oscillator which divides by 10 to give pulses for triggering another blocking oscillator whose output is a train of pulses of about 10 microsec duration also at a repetition frequency of 10 kc/s. These pulses, which are called 'gating pulses', are sent into a lumped-constant electrical delay line whose effective length is 100 microsec. At each tapping on the delay line a gating pulse is available to operate a crystal-diode coincidence gate. The gating pulses open gates 1–5 and close gate 6 at appropriate times.

To form the trains of timing pulses, five separate outputs are taken from the delay line and each is fed to one of 5 gates; thus each gate can be opened once every 100 microsec at a predetermined time which is arranged to overlap the time of a narrow pulse from the 100 kc/s blocking oscillator applied to the other input to the gates, and the output of each gate is therefore a narrow pulse selected from the 100 kc/s pulse train. The five trains of timing pulses in a suitable order are used to time the transitions of the output-pulse generators. It should be noted that the time stability of the pulses at the output of the gates is dependent only on that of the 100 kc/s pulse train from the 100 kc/s blocking oscillator. Variations of timing in the divider, the gating-pulse blocking oscillator and

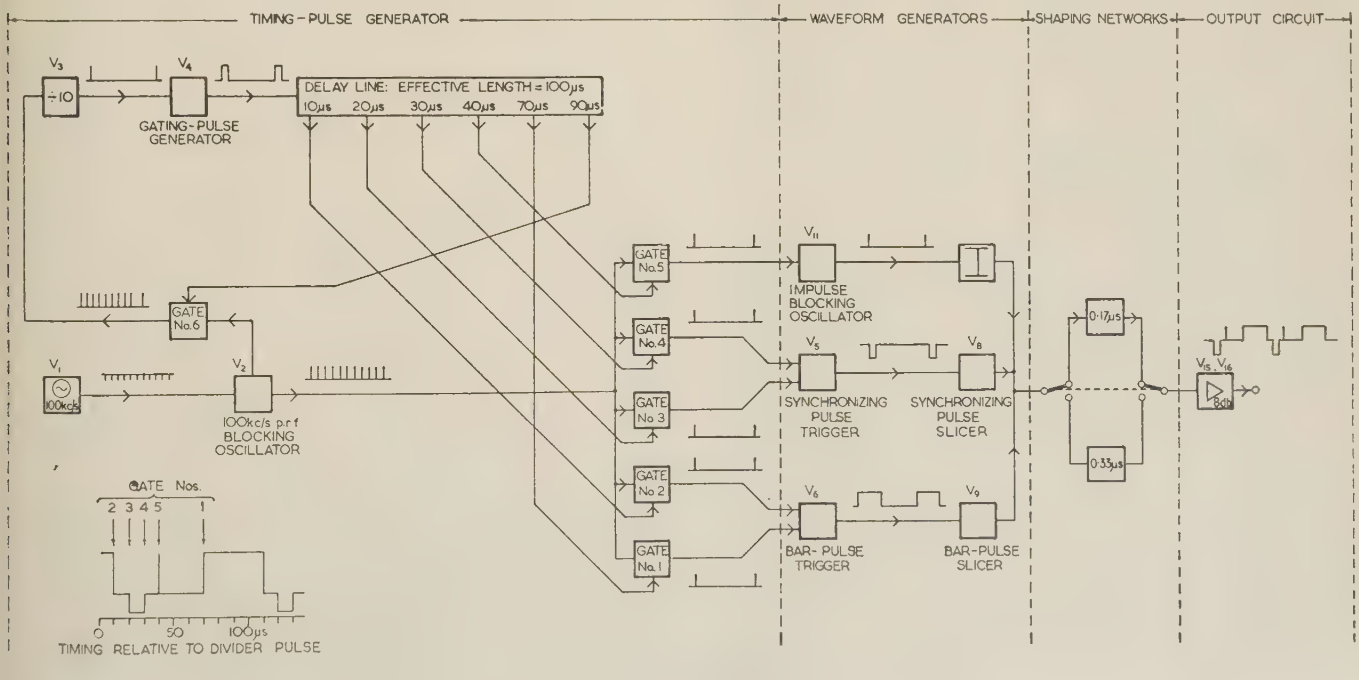


Fig. 2.—Block diagram.

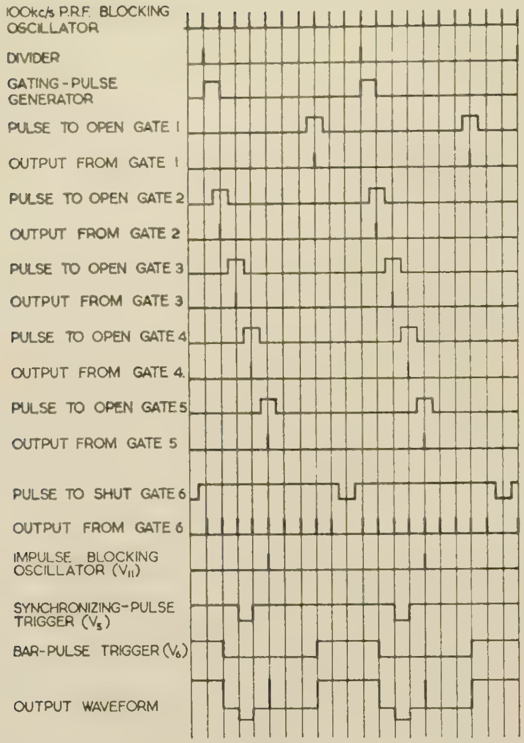


Fig. 3.—Timing chart.

the delay line will not affect any gate output, provided they are not so great as to prevent the coincidence of the gating pulse and the appropriate narrow pulse at the gate.

The outputs of gates 3 and 4 are 10 microsec apart, and they alternately change the state of a bistable trigger whose output is a negative-polarity pulse of 10 microsec duration, which is fed to a slicer, giving the synchronization-pulse component of the output waveform. The output of gate 5 occurs 10 microsec after that of gate 4, and triggers an 'impulse' blocking oscillator which provides the driving pulse for the sine-squared pulse component of the output waveform. The output of gate 1 occurs 30 microsec after that of gate 4, and the outputs of gates 1 and 2 are spaced 40 microsec apart. These pulses are used alternately to change the state of a bistable trigger whose output is a 40 microsec positive-polarity pulse which is fed to a slicer, giving the 'bar' component of the output waveform.

The three component pulse trains are then combined in a sine-squared shaping network in such a way as to give the required output waveform and at the same time to meet the requirements for the output voltage and impedance which were stated in the preceding Section. Experience has shown that, when the output is taken from the sine-squared shaping network,[2] a masking pad of at least 7 dB, and preferably about 18 dB, loss is required to avoid distortion of the pulse shape when the generator is connected to networks whose impedance varies with frequency. This presents no great difficulty when only a sine-squared pulse is required, but when the bar and synchronizing-pulse waveforms are also required, and where the rise time of the bar waveform must be very short, it becomes difficult and extravagant in power supplies to use this method. The bar and synchronizing-pulse trains are therefore connected to the input of the appropriate pulse-shaping network from high-impedance low-level circuits and the combined output from the network is connected to a high-quality video amplifier of about 8 dB gain. The amplifier provides the required 75-ohm output impedance and gives the necessary isolation between the load and the sine-squared shaping network.

(3) CIRCUIT DETAILS

(3.1) Timing Generator

(3.1.1) Generation of Timing Pulses.

The 100 kc/s train of timing pulses controls directly the timing of the output waveform and must therefore be as free from jitter as possible. This requires an oscillator of good short-term phase stability, the generation of pulses with very short rise-times, and the choice of circuits which minimize the possibility of the timing or amplitude of the pulses being modified by residual ripple in the power supplies. Fig. 4 shows the circuit chosen, which uses

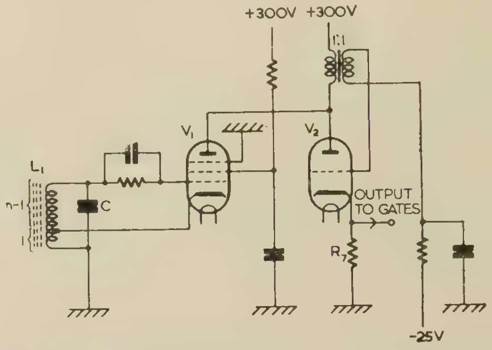


Fig. 4.—Timing generator showing 100 kc/s oscillator (V_1) and 100 kc/s blocking oscillator (V_2).

a cathode-tap Hartley oscillator, V_1, arranged to trigger a blocking oscillator, V_2. The oscillator operates in the class-C mode and the bias across the grid leak is such that the valve is cut off during the greater part of the cycle; the current therefore flows only in short pulses which are derived from the tips of a sine wave. The duration of the current pulse should be as short as is practicable. An approximate analysis of the circuit shows the duration to be a function of $n^2/(n-1)R_D$, where n is determined by the position of the tapping on the coil and R_D is the effective dynamic resistance of the tuned circuit, including the effects of grid-current damping. Frequency-stability requirements set a practical upper limit on R_D; $n^2/(n-1)$ is a minimum with $n = 2$, i.e. with a centre-tapped coil. The use of this value of n in the present case leads to an excessive voltage swing on the cathode, which would considerably reduce the magnitude of the output pulse because an unlimited screen-voltage supply is not available. A value of $n = 4$ was therefore chosen, because this reduces the peak cathode voltage to one-third of the value when $n = 2$ and increases the pulse width by only about 12%. With the values chosen the amplitude of the sine wave across the tuned circuit is 300 volts d.a.p. giving a cathode voltage of 75 volts d.a.p. and an anode-current pulse of approximately 1 microsec duration (measured at its base).

A high Q-factor in the oscillator tuned circuit is an essential for good phase stability and it is worth remarking that the large signal across the ferrite-cored inductor L_1 reduces its Q-factor to 60% of the small-signal value of 260.

The anode of the oscillator valve is connected to the anode of a blocking oscillator, V_2, which is normally held at cut-off. A current pulse received from the oscillator causes the blocking-oscillator anode voltage to fall, with a consequent rise in grid voltage, owing to the 1 : 1 coupling transformer. When the grid voltage has reached the point where the loop gain of the circuit is unity, regeneration takes place and a short current pulse flows through the valve.[2] A pulse of some 50 volts amplitude with a

half-amplitude duration of 0·2 microsec appears across R_7, and the train of pulses so formed is used to control the timing of the output waveform after time selection by the gates.

(3.1.2) Generation of Gating Pulses.

As indicated in Section 2, the gating pulses, which are of 10 kc/s repetition frequency and wide enough to overlap one pulse from the 100 kc/s blocking oscillator, are obtained by using a frequency divider driving a gating-pulse blocking oscillator. The circuit arrangement is shown in Fig. 5.

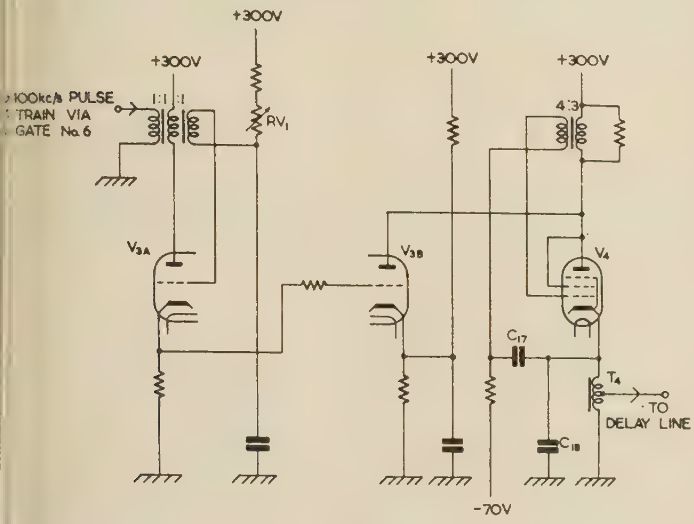


Fig. 5.—Timing generator showing frequency divider (V_{3A}), buffer (V_{3B}) and 10 kc/s gating pulse generator (V_4).

The divider is a free-running blocking oscillator, V_{3A}, with a natural period of about 100 microsec, synchronized by every tenth pulse from the 100 kc/s blocking oscillator. The grid time-constant of the divider is adjustable by means of the preset variable resistor RV_1, to allow for differences in valves and to enable the correct division ratio to be set up. To stabilize the divider against deterioration of the valve with age and also against reduction of supply voltages (the two most common causes of change of division ratio), every ninth pulse of the input from the 100 kc/s blocking oscillator is suppressed by gate 6 (see Section 3.1.3). Any tendency for the divider to lock on the ninth pulse is therefore eliminated, and a greater margin of stability is achieved.

The gating-pulse blocking oscillator, V_4, is triggered through a buffer stage, V_{3B}, by pulses from the divider. The design of the blocking oscillator was based on Benjamin's[12] 'sinh-sin' mode and the nominal pulse width is 10 microsec. The grid capacitor, C_{17}, is connected to the cathode to improve the stability of pulse width against supply-voltage changes. An auto-transformer, T_4, is used to deliver the maximum pulse amplitude (about 40 volts) into the delay line, and the capacitor C_{18} reduces high-frequency ringing caused by the auto-transformer.

A total delay of 90 microsec is required in the delay line. To minimize the volume and cost of the delay line, the largest practicable delay per section (with consequent narrow bandwidth) must be used in conjunction with the widest practicable gating pulse. The use of wide-tolerance paper capacitors and low-Q-factor inductors is also dictated by considerations of cost, and the consequent distortion of the pulse in its progress down the line necessitates some compromise in the design bandwidth of the line and in the pulse width. These considerations led to the adoption of 10 microsec as the pulse width and 200 kc/s as the nominal cut-off frequency of the line. The conventional type of lumped-constant line with mutually coupled coils is used and the nominal impedance is 100 ohms.

A further reduction in volume is obtained by using a line of 50 microsec length with a short-circuit termination. Thus, by using the reflected pulse, delays up to the longest required, i.e. 90 microsec, can be achieved. The delay of 2 microsec per section gives a sufficient number of tappings on the line to ensure that it is always possible to overlap any required 100 kc/s pulse with a gating pulse. Because of the high dissipation of the line, owing to the inexpensive components used, the pulse attenuation is approximately 0·1 dB/microsec, and at the 40 and 70 microsec tappings step-up transformers are used to ensure that the gating pulses at these times are of sufficient amplitude to open gates 5 and 1 (see Section 3.1.3) under all circumstances. The transformer on the 70 microsec tapping also inverts the pulse from the line, since the reflected (negative polarity) pulse is required, and the gates are opened only by positive pulses.

(3.1.3) Gate Circuit.

Each gate acts as a linear on/off switch[10] and the output pulse is a replica of the selected input pulse. Fig. 6 shows the circuit

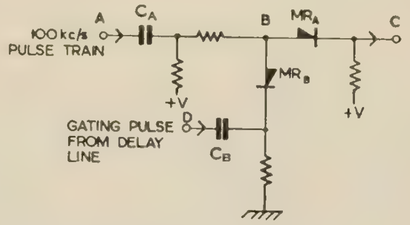


Fig. 6.—Diode gate circuit.

arrangement in which the switching elements are point-contact germanium diodes, MR_A and MR_B. The train of 50-volt positive pulses of approximately 0·2 microsec half-amplitude duration from the 100 kc/s blocking oscillator constitutes the signal to the gate at terminal A. The bias V is chosen so that, with no positive gating pulse at terminal D but with a signal input at terminal A, MR_B conducts and MR_A is cut off. There is thus a low-impedance path presented to the signal via MR_B, C_B and the delay line to earth, and the signal appearing at terminal B is so small that a negligible amplitude appears at terminal C, since the back resistance of MR_B is greater than 1 megohm and the load resistance is 4·7 kilohms. However, when a positive gating pulse from the delay line appears at terminal D, the cathode potential of MR_A rises and this diode is cut off. This reduces the current drawn from source V, the potential at terminal B rises, and MR_A is then held ready to conduct, and when the next signal pulse occurs, it appears at the output terminal C. The output waveform is free from 'pedestal' because of the presence of MR_A. For maximum output the cathode potential of MR_B must rise by an amount at least as great as the amplitude of the selected signal pulse which appears at terminal B.

Gate 6, which is introduced to improve the stability of the divider, is similar in circuit arrangement to the other gates, except that the capacitor C_A is omitted and the bias voltage is zero. A positive bias is developed at the junction of C_B and MR_B by rectification of the timing pulses applied to terminal A. This bias allows the timing pulses to be transmitted to terminal C with reasonably small attenuation in the absence of a negative gating pulse at terminal D. A negative pulse at terminal D causes conduction of MR_B and non-conduction of MR_A and so suppresses the appropriate timing pulse.

(3.2) Generation of Output Pulses

(3.2.1) Synchronizing and Bar Pulses.

Requirements for the bar pulse are that the leading and trailing edges should have rise times of 0·07 microsec or less and that the horizontals should be flat to within ±0·2%. The synchronizing pulse should be well shaped. In practice it was found convenient to generate both pulses in a similar manner and Fig. 7

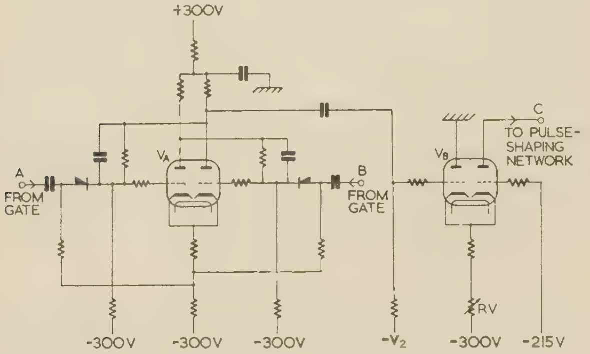


Fig. 7.—Waveform generator for bar or synchronizing pulse.
V_A Trigger.
V_B Amplitude slicer.

shows the circuit used for each. V_A is a bistable cathode-coupled trigger[10] with signal-injection diodes. The circuit is symmetrical and the use of a high-value resistor in the common cathode lead ensures reliable triggering with variations in valves and supply voltages.[13] A current of 10 mA is switched regeneratively from one half of the valve to the other when a narrow positive pulse is received from the gate connected to terminal A or B. The output pulse amplitude is about 30 volts peak, the duration of the pulse being determined by the interval between the narrow pulses connected to terminals A and B. This waveform is fed to V_B, and the bias voltage, V_2, is chosen to dispose the waveform symmetrically about −215 volts. V_B acts as an amplitude slicer, and a clean pulse with a rise time less than 0·04 microsec appears at the output terminal C. R_V controls the current in the valve and is adjusted to give the appropriate amplitude to each pulse, namely 0·15 volt for the synchronizing pulse and 0·35 volt for the bar pulse.

(3.2.2) Sine-Squared Pulse.

The sine-squared pulse is generated by feeding an impulse from a blocking oscillator into a shaping network (see Section 3.3). The blocking oscillator, shown in Fig. 8, is triggered by the pulses from gate 5 applied via the buffer valve V_{10}. On receipt of the trigger pulse the blocking-oscillator valve V_{11} generates a pulse of about 25 volts amplitude and 35 millimicrosec half-amplitude duration in the cathode load. The half-amplitude duration is determined mainly by C_{83} and by the design of the transformer T_7. This mode of operation of the blocking oscillator has been described in some detail.[2] A conventional shunt voltage-stabilizer, V_{12}, is included to stabilize the pulse amplitude against changes in the h.t. supply. RV_4 permits manual control of the amplitude of the sine-squared pulse to be obtained with a range of approximately 6 dB, so that a pulse-to-bar ratio of unity in the output waveform may be achieved with normal variations in the characteristics of the valves.

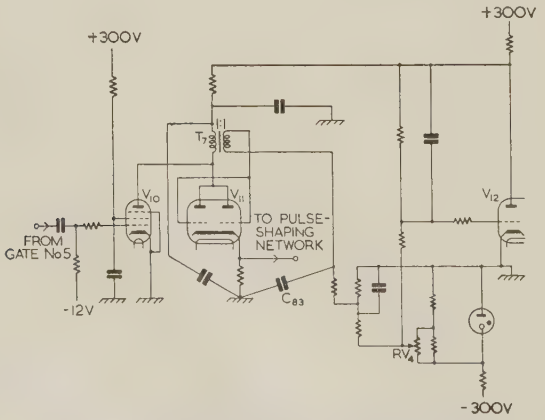


Fig. 8.—Impulse blocking oscillator and stabilizer.

(3.3) Pulse Mixing and Shaping

The heart of the generator is the sine-squared pulse-shaping network, the design of which is derived from Solution 3 in the paper by Thomson.[11] This network gives a slightly better approximation to the true sine-squared shape than that originally used.[2] The network is shown in Fig. 9 and its calculated impulse

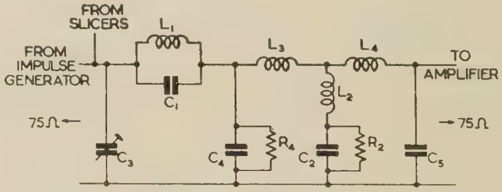


Fig. 9.—Sine-squared pulse-shaping network.

Component	Value: For half-amplitude duration of 1/6 microsec	Value: For half-amplitude duration of 1/3 microsec	Tolerance (%)	Q-factor†
L_1	1·580 µH	3·159 µH	±1	⩾ 70
L_2	0·308 µH	0·617 µH	*	⩾ 50
L_3	3·091 µH	6·182 µH	±1	⩾100
L_4	3·035 µH	6·069 µH	±1	⩾100
C_1	79·22 pF	158·4 pF	±2	
C_2	2 168 pF	4 335 pF	±0·5	
C_3	75·92 pF	151·8 pF	±2	
C_4	566·4 pF	1 133 pF	±0·5	
C_5	166·4 pF	332·9 pF	±2	
R_2	1 300 ohms	1 300 ohms	±5	
R_4	5 100 ohms	5 100 ohms	±5	

* L_2 should be adjusted to make the insertion loss a maximum at 6·156 Mc/s for the 1/6 microsec network, and at 3·078 Mc/s for the 1/3 microsec network.
† The Q-factors should be measured at 4 Mc/s for the 1/6 microsec network, and at 2 Mc/s for the 1/3 microsec network.
Note.—An allowance for stray capacitance should be made in the value of any capacitor if the total capacitance would otherwise exceed the given limit.

and step responses are shown in Fig. 1. The Q-factors of the coils and the tolerances on components were chosen after experiment on a low-frequency model of the network. These experiments also showed that the 'best' pulse shape (i.e. that nearest the theoretical impulse response of the network) was obtained by making the network uniformly dissipative. To do this the Q-factor of each capacitor must be degraded to equal the harmonic mean of those of the coils; R_2 and R_4 perform this function in the network. The values of the other capacitors are such that the addition of the appropriate resistors has negligible effect on pulse shape, and so they are omitted.

The network is designed to work between resistive terminations which should be 75 ohms ± 1%. To avoid excessive attenuation of the bar and synchronizing-pulse components of the waveform, these are fed directly from the anodes of the slicers to the input of the network, and the output of the network is connected, through a small masking pad (approximately 2·5 dB)

and the switch, to the amplifier, whose input impedance is made 75 ohms $\pm 1\%$. A further pad is used at the input of the network to reduce the level of the impulse and give the correct amplitude of sine-squared pulse at the output of the network. This pad is in two parts, one on the blocking-oscillator side of the switch and the other on the network side, to reduce level differences between the switch wafers. The part associated with the network has approximately 6 dB greater loss for the *T*-pulse network than for the $2T$ network, to allow for the difference in the shaping loss[2] of the networks.

The anode impedance of the two slicer valves, together with the impedance of the connecting cable, is very nearly purely capacitive. The effect of this is allowed for by an appropriate reduction of C_3 in the network, allowance being made for wiring variations by adjustment of a trimmer.

(3.4) Output Amplifier

The gain required in the amplifier is about 8 dB, to give a 1-volt d.a.p. waveform across 75 ohms at its output. The circuit shown in Fig. 10 has two triode stages and a cathode-follower output with overall negative feedback. Triodes have been found to be more suitable than pentodes in a low-gain video feedback amplifier. C_{65} controls the high-frequency response, and it is adjusted so that there is no change in the T pulse-to-bar amplitude ratio from input to output of the amplifier. The input potential divider enables the gain to be set to give a 1-volt output signal. To allow for the manufacturing tolerance in the value of the variable resistor, R_{109} is selected so that, in combination with the variable resistor and R_{110}, the input impedance is 75 ohms $\pm 1\%$. The amplifier gain/frequency characteristic is flat to within 0·01 dB up to 11 Mc/s. The return loss of the output impedance against 75 ohms is better than 30 dB. The d.c. component at the output of the amplifier is about 2·25 volts across a 75-ohm load.

Because the h.t. supply is unregulated, a valve smoothing circuit, V_{14}, is used in the h.t. feed to the amplifier to reduce hum and also 'bumping' due to fluctuations in the mains supply. The circuit is based on the mutual-conductance bridge and has been known for a long time.[14] The arrangement actually used, shown in Fig. 10, has a large amount of cathode degeneration,[15] which improves the stability of interference rejection against variations in valve characteristics—by a factor of 14 in the present case. The degeneration also increases the grid time-constant by a factor of 6 which, for a modest grid capacitance, enables the circuit to deal with the slowest fluctations which are likely to cause trouble at the amplifier output. No adjustment has been found necessary to deal with initial tolerances or ageing of valves.

(3.5) Power Supplies

The circuits, with the exceptions already mentioned, are designed to function correctly with a reasonable variation of h.t. supply (+5 to −10%) and a moderate amount of hum. This permits the use of an unregulated power supply, with consequent economy in cost, bulk and weight. One mains transformer feeds both positive and negative h.t. lines and the heaters. The h.t. power supplies are conventional full-wave valve-rectifier circuits with capacitive input smoothing filters. Since the negative line takes less current than the positive line, the transformer is tapped down to give a nominal 300 volts on the negative line under load. The current consumption on the +300-volt line is about 145 mA, and on the −300-volt line, about 70 mA. The ripple on the lines is approximately 0·15-volt d.a.p., which is sufficiently small to have no effect on the waveform up to the input of the amplifier. The amplifier h.t. smoothing circuit reduces the ripple to negligible proportions, so that the waveform at the output of the generator is substantially free from ripple.

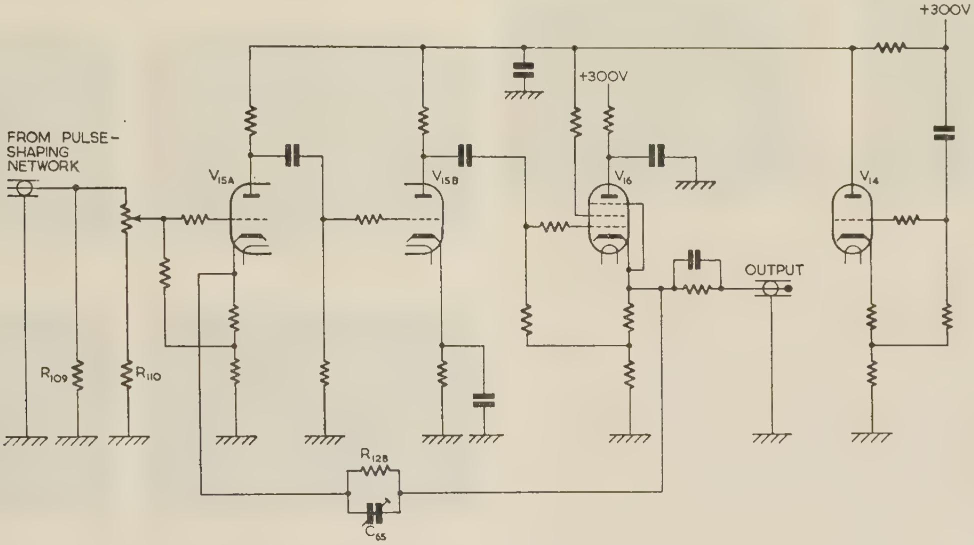


Fig. 10.—Output amplifier.

(4) PERFORMANCE

Photographs of the output waveforms are shown in Fig. 11. In addition to showing the complete $2T$ waveform the photographs show on expanded time scales the T and $2T$ sine-squared pulses, details of the $2T$ bar edges and a superimposed display of the $2T$ pulse and bar edge. The most sensitive indication of the correctness of the sine-squared pulse shape is given by the shape and magnitude of the first (negative) and second (positive) overshoots which follow the main lobe of the pulse and which should be similar to the calculated response shown in Fig. 1. Provided that the components in the shaping networks are set up to within the tolerances specified in Fig. 9 and the terminating impedances are correct, the pulse shape at the output of the unit is such that the first overshoot does not differ by more than $\pm 0{\cdot}5\%$ from its nominal value of about 0·9%, and the second overshoot by more than $\pm 0{\cdot}25\%$ from its nominal value of about 0·4%. The flatness of the bar top is better than 0·2%.

The amplitude of the output waveform changes by less than

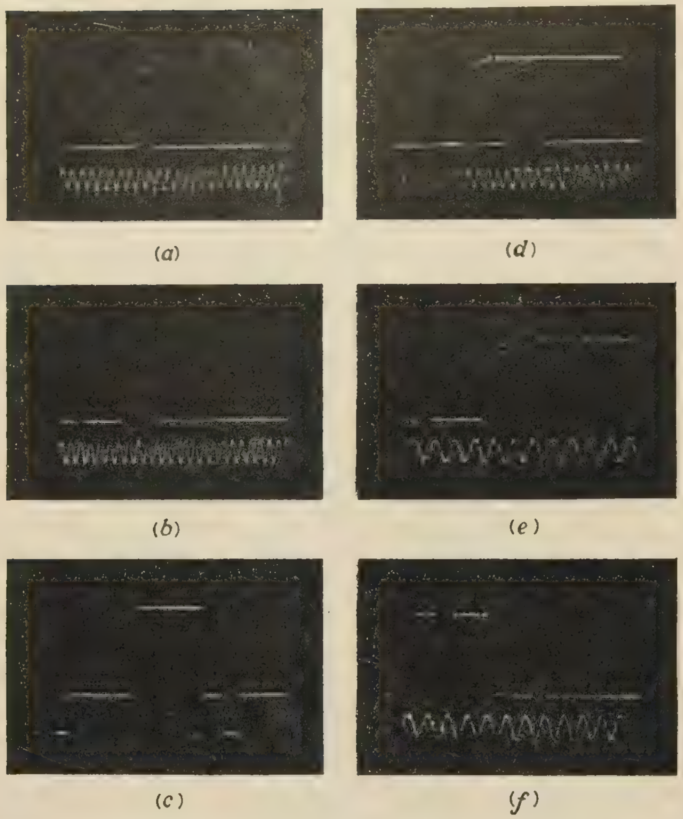


Fig. 11.—Pulse-and-bar waveforms for testing television links.
Timing waves: (*a*), (*b*) and (*d*) are 6 Mc/s; (*e*) and (*f*) are 1 Mc/s.

$\pm 2\%$ and the bar/pulse amplitude ratio changes by less than $\pm 1\%$ when the mains supply voltage is varied by $\pm 5\%$ from the nominal.

The stability of the timing of the waveform depends on the frequency stability of the 100 kc/s master oscillator. Tests on a typical production unit showed that the frequency change was less than $0 \cdot 1\%$ for an ambient temperature change of 20–45° C, and at any temperature in this range the frequency did not change by more than $\pm 0 \cdot 05\%$ for a mains variation of $\pm 15\%$ from the nominal. During these tests the unit was enclosed in a dust cover, which gives a greater internal temperature rise than the better-ventilated portable carrying case, and it is worth noting that all the circuits continued to operate satisfactorily even when the ambient temperature range was extended to 0–50° C.

The jitter of any part of the waveform relative to the synchronizing pulse is too small to measure, but is certainly less than 1 millimicrosec.

More than 100 generators of this type have been manufactured without any special difficulties, and the results show that the circuits are reproducible and that the output waveform does not vary significantly from unit to unit.

(5) FURTHER DEVELOPMENTS

The pulse-and-bar waveform generator described in the previous Sections can be modified to test links for television systems on other standards by suitably scaling the frequency of the timing generator and/or the sine-squared shaping networks. It is also possible to use pulse-and-bar waveforms in an analogous manner for additional tests on links which may carry colour television signals. A brief description will now be given of two such modified generators which have been constructed.

(5.1) Generator for 625-Line 5 Mc/s System

This is a straightforward conversion for a 5 Mc/s system, where the line repetition frequency is approximately 16 kc/s and the T and $2T$ pulses are of $0 \cdot 1$ and $0 \cdot 2$ microsec half-amplitude duration respectively. This generator differs from that detailed in Sections 1–4 in that the main oscillator (Fig. 4) has a frequency of 160 kc/s, obtained by the appropriate change in L_1 and C; the divider (Fig. 5) has a repetition frequency of 16 kc/s, obtained by alteration of the time-constant of the grid of V_{3A}; the delay-line has an effective length of $62 \cdot 5$ microsec; the impulse feeding the sine-squared networks has a half-amplitude duration of 20 millimicrosec, obtained by redesign of T_7 and change of C_{83} (Fig. 8), and the components in the sine-squared shaping networks (Fig. 9) are scaled in value to give the $0 \cdot 1$ and $0 \cdot 2$ microsec sine-squared responses. The waveforms obtained are, of course, similar to those shown in Fig. 11, with the appropriate time-scale changes.

(5.2) The Testing of Links for Colour Television

Colour-television systems in which one or more sub-carriers are included within the normal monochrome video band naturally

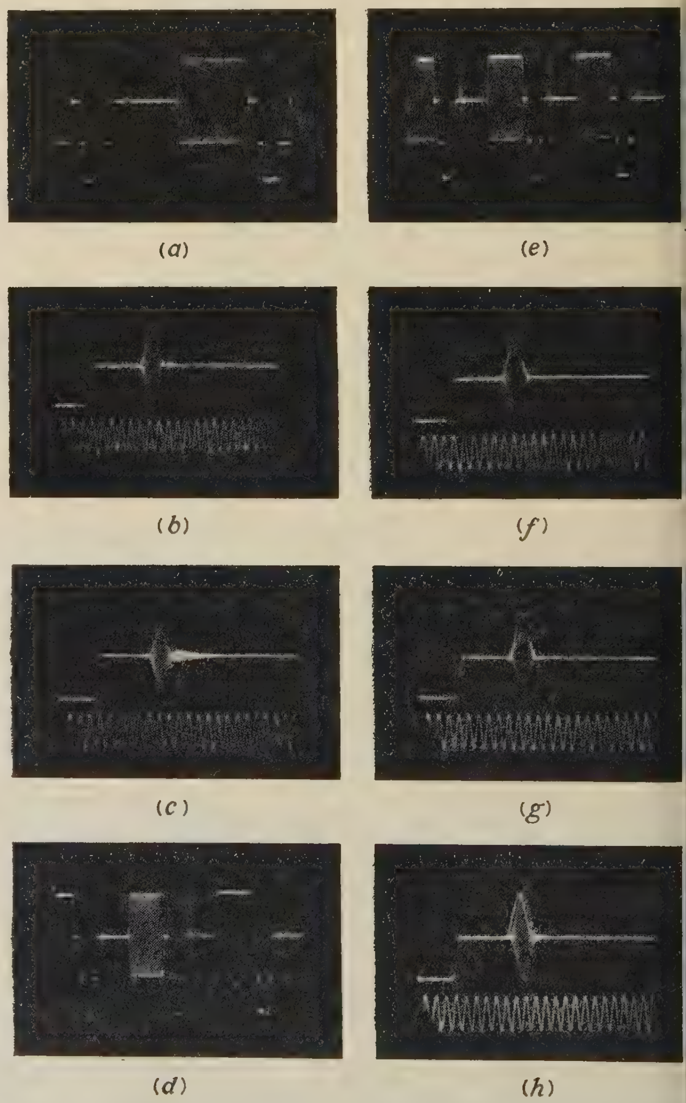


Fig. 12.—Pulse-and-bar waveforms for testing colour-television links
Timing waves: 1 Mc/s.

impose additional requirements upon the performance of the links. Although the response of a link to the normal pulse-and-bar test signal includes information on distortions occurring in the vicinity of the sub-carrier frequencies, this information is not in a form which can easily be interpreted in terms of distortions of the sub-carrier signals. However, a simple extension of the concepts of monochrome waveform testing enables the important linear distortions affecting colour signals to be measured directly. In this Section are given, to illustrate the principles, some brief details of test signals and methods which would be suitable for use with a colour television system of the N.T.S.C. type.

(5.2.1) Test Signals.

The proposed additional test signals are closely analogous to the normal pulse-and-bar signals; the essential principle is that the sub-carrier is amplitude-modulated (double-sideband) by pulse-and-bar combinations whose T and $2T$ half-amplitude durations and rise times are chosen to suit the nominal bandwidth(s) of the chrominance channel. Fig. 12(a) shows how the modulated sub-carrier is raised on a pedestal between normal line-synchronizing pulses, and Fig. 12(b) shows the sine-squared pulse portion of the waveform on an expanded time scale. By operating a switch on the generator, the signal can be changed to that shown in Fig. 12(d), where alternate line periods contain the modulated sub-carrier waveform and the modulating (or 'video') waveform respectively. This arrangement enables the two waveforms to be superimposed on an oscillograph display if the time-base is triggered from every line-synchronizing pulse. In the absence of distortion, the upper half-envelopes of the modulated sub-carrier elements are in precise registration in both time and amplitude with the corresponding elements of the video waveform, as shown in Figs. 12(e) and 12(f).

(5.2.2) Luminance-Channel Waveform Distortion.

Because the effects of waveform distortion on the luminance component of the N.T.S.C.-type signal are essentially the same as those mentioned in Section 1 for the monochrome signal, the same methods of measurement can be used. Thus the normal pulse-and-bar test signals, with values of T and $2T$ appropriate for the monochrome system, may be adopted for measuring the luminance-channel waveform distortion over the video spectrum from the line-repetition frequency upwards.

(5.2.3) Chrominance-Channel Waveform Distortion.

Modulated sub-carrier pulse-and-bar signals of the type shown in Figs. 12(a) and 12(b) are suitable for the measurement of waveform distortion in the chrominance channel. It is not necessary to demodulate this signal after transmission over a link, because it can be viewed directly on an oscillograph and because measurement of the distortion suffered by the sub-carrier envelope is considered to be an effective practical test. Rating factors can be applied in a manner similar to that outlined in Section 1. Fig. 12(c) shows an example of the effect of distortion on the sine-squared pulse portion of the waveform; it will be seen that the pulse is reduced in amplitude and is followed by a substantial 'tail'.

(5.2.4) Luminance-Chrominance Amplitude Inequality.

Any disturbance of the relative amplitudes of the luminance and chrominance signals causes colours to be under- or over-saturated with respect to white. To test for this form of distortion, the N.T.S.C. recommends, in effect, a comparison of the steady-state gain at the sub-carrier frequency with that at the line-repetition frequency. However, the amplitude inequality can be tested more directly with the test signal shown in Fig. 12(d) by comparing, in the superimposed display, the single-peak amplitude of one of the modulated sub-carrier elements with the amplitude of the corresponding video element. Fig. 12(g) shows an example of the effect of amplitude inequality on the sine-squared pulse portions of the waveform, but in practice it would be best to make the comparison at the mid-point of the bar portions, because this is equivalent to a comparison of large-area white and coloured portions of the picture.

(5.2.5) Luminance-Chrominance Delay Inequality.

Any difference between the transmission delays of the luminance and chrominance signals causes a registration error in the colour picture. To test for this form of distortion the N.T.S.C.

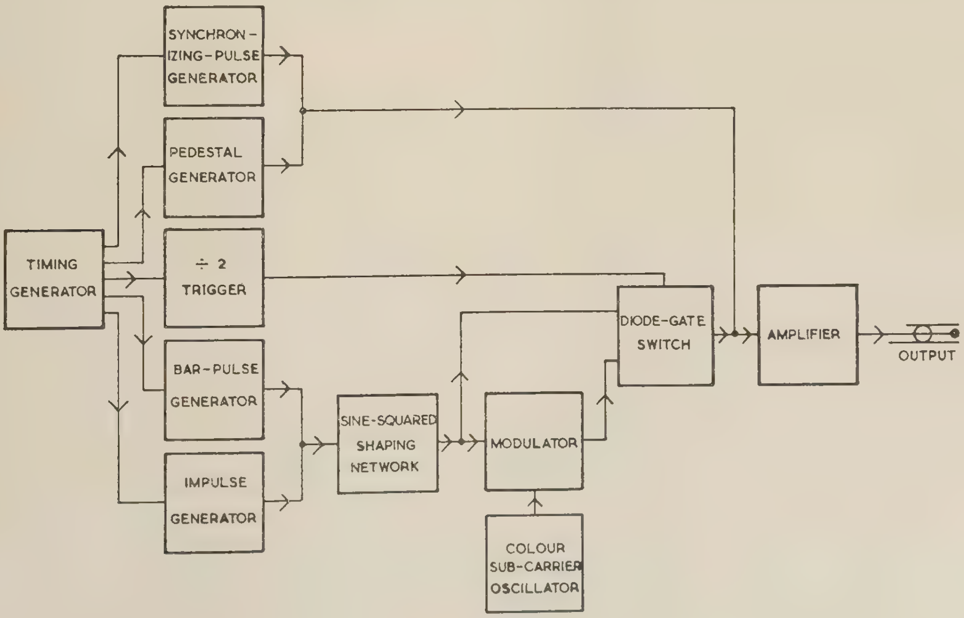


Fig. 13.—Block diagram of video and modulated sub-carrier pulse-and-bar generator.

recommends, in effect, a comparison of the steady-state envelope delay at the sub-carrier frequency with the average envelope delay of a band (50–200 kc/s) near the line-repetition frequency. Again, as in the case of amplitude inequality, the delay inequality can be tested more directly with the test signal shown in Fig. 12(*d*). The method here proposed is to make a direct comparison of the timing, relative to the synchronizing pulse, of corresponding features of the modulated sub-carrier and video elements. The feature chosen for comparison could be either the peak (or, more precisely, the mid-half-amplitude point) of the sine-squared pulse or the half-amplitude point of one of the bar transitions. Fig. 12(*h*) shows an example of the effect of delay inequality on the sine-squared pulse portions of the waveforms.

It will be appreciated that the waveform method not only obviates the need for the difficult measurement of average envelope delay near the line-repetition frequency, but gives a direct test of what should be the chief point of interest, namely the time coincidence of associated luminance and chrominance information.

(5.2.6) Test-Signal Generator.

The waveforms shown in Fig. 12 were produced by an experimental generator made, as indicated in the block schematic in Fig. 13, by modifying a normal pulse-and-bar generator and adding certain units. The shaping networks were changed to give T and $2T$ half-amplitude durations and rise times of 4/3 and 8/3 microsec respectively to suit a possible 405-line scaled-down N.T.S.C. system with a sub-carrier frequency of 2·66 Mc/s.

(6) ACKNOWLEDGMENTS

Thanks are due to our colleagues—Mr. H. J. Orchard, for the design of the sine-squared pulse-shaping networks, and Mr. H. S. Hale, for the design of the output amplifier. We are indebted to Dr. N. W. Lewis for his continued interest in the work and for his help in the preparation of the paper, particularly in respect of Section 5.2, for which he is jointly responsible.

Acknowledgment is made to the Engineer-in-Chief of the Post Office for permission to make use of the information contained in the paper.

(7) REFERENCES

(1) Lewis, N. W.: 'Waveform Responses of Television Links', *Proceedings I.E.E.*, Paper No. 1688 R, July, 1954 (**101**, Part III, p. 258).

(2) Macdiarmid, I. F.: 'A Testing Pulse for Television Links', *ibid.*, Paper No. 1267 R, April, 1952 (**99**, Part IIIA, p. 436).

(3) Hunt, G. J., and Elliot, E. W.: 'An Introduction to the Sine-Squared Pulse', *Journal of the Television Society*, 1953, **7**, p. 49.

(4) Kennedy, R. C.: 'Sine-Squared Pulses Test Color-TV Systems', *Electronics*, December, 1954, **27**, p. 138.

(5) Cooper, V. J.: 'High-Power Television-Transmitter Technique, with particular reference to the Transmitter at Holme Moss', *Proceedings I.E.E.*, Paper No. 1297 R, April, 1952 (**99**, Part IIIA, p. 231).

(6) Cooper, V. J.: 'The Transient Testing of Television Apparatus', *Marconi Review*, 1953, **16**, p. 1.

(7) Clayton, R. J., Espley, D. C., Griffiths, G. W. S., and Pinkham, J. M. C.: 'The London–Birmingham Television Radio-Relay Link', *Proceedings I.E.E.*, Paper No. 1113 R, February, 1951 (**98**, Part I, p. 204).

(8) Kilvington, T., Laver, F. J. M., and Stanesby, H.: 'The London–Birmingham Television-Cable System', *ibid.*, Paper No. 1187 R, October, 1951 (**99**, Part I, p. 44).

(9) Grayson, H., McLeod, T. S., Dunkley, R. A. G., and Dawson, G.: 'Circuit Techniques in Frequency-Modulated Microwave Links', *ibid.*, Paper No. 1291 R, March 1952 (**99**, Part III, p. 256).

(10) Chance, B., *et al.*: 'Waveforms', M.I.T. Radiation Laboratory Series (McGraw-Hill, 1949), Chapters 5 and 10.

(11) Thomson, W. E.: 'The Synthesis of a Network to have a Sine-Squared Impulse Response', *Proceedings I.E.E.*, Paper No. 1388 R, November, 1952 (**99**, Part III, p. 373).

(12) Benjamin, R.: 'Blocking Oscillators', *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 1159.

(13) Paivinen, J. O., and Auerbach, I. L.: 'Design of Triode Flip-Flops for Long-Term Stability', *Transactions of the Institute of Radio Engineers*, EC–2, 1953, p. 14.

(14) 'Valve-Operated Smoothing Circuit', *Wireless World*, November, 1939, **46**, p. 28.

(15) Sowerby, J. McG.: 'Shunt Voltage Stabilizer', *ibid.*, June 1948, **54**, p. 200.

:21.376.53 : 621.395.4

The Institution of Electrical Engineers
Paper No. 2474 R
Dec. 1957

EFFICIENCY AND RECIPROCITY IN PULSE-AMPLITUDE MODULATION: PART 1—PRINCIPLES

By K. W. CATTERMOLE, B.Sc.

The paper was first received 2*nd May, and in revised form* 21*st October*, 1957. *It was published in December*, 1957, *and was read before the* Radio and Telecommunication Section 19*th March*, 1958.)

SUMMARY

The paper forms the first part of two papers on efficiency and ciprocity in pulse-amplitude modulation. A method of converting low-frequency signal into a modulated pulse train and back again, ith low power loss, provides multiplex communication on a 2-wire sis, without amplifiers. In principle, lossless 2-way channels with bandwidth of half the sampling rate are feasible. The paper furnishes theoretical study of the transmission properties of such channels.

(1) INTRODUCTION

Modulation methods are usually judged according to the fficiency with which they convey signal information rather than gnal power. This is defensible in communication systems here the primary problem is to combat noise or interference, nd the provision of amplifiers is, technically and economically, nly a small part of the whole. In the application of multiplex ethods to telephone switching, however, the signal is not egraded by a long, noisy and band-limited path: the primary quirements are economy and reliability, the need for which is ven greater than in most other communication systems because f the large quantities of apparatus used. A method for conying signal power through the entire path with low loss, transorming from audio-frequency signals to a modulated pulse train nd back again to audio in an efficient manner, permits subantial reduction in apparatus. The method which has evolved uring this study also offers improvements in crosstalk margin, ain stability, and, when noisy devices such as transistors are sed, in noise level.

(2) PHYSICAL PRINCIPLES

(2.1) Efficiency

The processes of pulse modulation and demodulation are sually performed in a very inefficient manner, in that (*a*) since sample of short duration is drawn from the source, the available ower of the source must be greatly in excess of the pulse power btained, and (*b*) since only the signal-frequency component of e received pulse is utilized in the receiver, only a small fraction f the pulse power is reconverted into signal. Communication ystems therefore employ circuits which, whether they are simple mplifiers or more sophisticated devices, effectively give a very igh power gain. With valves, the performance can be obtained t the cost of more apparatus: with transistors, the performance ay be difficult to obtain at all.

The approach in the present study is to make the modulation nd demodulation processes inherently very efficient, so that little r no amplification is required in other parts of the system, i.e. continuous signal is to be converted into a modulated pulse ain of substantially the same mean power, and the pulse train ter reconverted to continuous signal. The principle is applible only to systems in which the energy of a pulse is proportional to the power of the modulating signal, and at present only pulse-amplitude modulation (p.a.m.) has been used.

In simple p.a.m. the modulator is effectively a sampling switch, as S in Fig. 1, which is periodically closed for a short time, say a

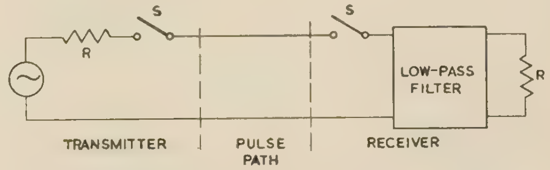


Fig. 1.—Simple p.a.m. transmission system.

fraction $1/A$ of the interval between successive closures. The mean power contained in the pulse train is at most a fraction $1/A$ of the available power from the audio source. The receiving filter extracts the audio component, whose power is at most $1/A$ of the mean power in the pulse train. Somewhere in the chain, therefore, a gain of A^2 must be provided. In practice, A will be in the region 50 to 100, for a 25-channel signal.

Efficient p.a.m. uses storage reactances which may be charged and discharged at different rates. In a modulator, the store is charged by a substantially smooth flow of power from the signal source, and discharged rapidly in the form of a short pulse. In a demodulator, the store is charged by a pulse of energy, and discharged to provide a substantially smooth flow of power. Thus, a modulator or demodulator consists essentially of a passive reactance network which may be connected by an electronic switch alternately to two circuits, a source and a load, with different time-constants in the two connections. In practice, the circuit with long time-constant is permanently connected, and that with short time-constant is periodically connected to obtain an impulsive charge or discharge, so that the basic circuit is as shown in Fig. 2.

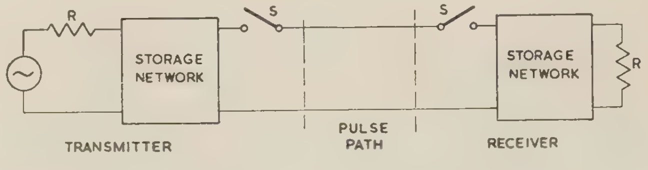


Fig. 2.—Storage p.a.m. transmission system.

The idea of reactance storage is, of course, an old one: it is the basis of delay-line pulse generators. However, the application to efficient communication is novel and promises to be very fruitful.

(2.2) Storage Networks and Transmission Performance

Detailed analysis of networks with periodic switch connection has revealed the ultimate limits to the transmission performance obtainable, a method of synthesizing storage networks to approach this ideal limit, and a method of calculating the overall

Mr. Cattermole is with Standard Telecommunication Laboratories Ltd.

transmission with any given storage network. The theory has been confirmed by a variety of experimental evidence, some of which is described in a companion paper.[13]

Before going into details, we summarize the results of the study. It is possible, in principle, to transmit signals within a bandwidth equal to half the sampling rate entirely without loss. The storage network required is then a delay-line section (or tuned circuit) chosen to provide a short square pulse (or half-cosine pulse) on discharge, which also constitutes the terminal capacitance of an asymmetrical low-pass filter with a suitable insertion-loss specification (Fig. 3). The transmission obtained

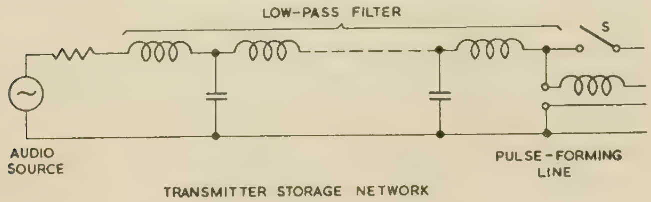


Fig. 3.—Storage network.

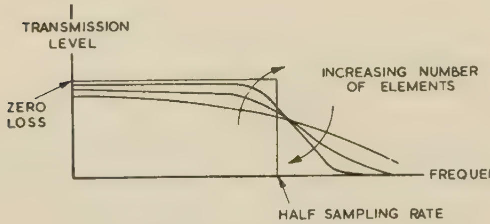


Fig. 4.—Transmission level as a function of frequency.

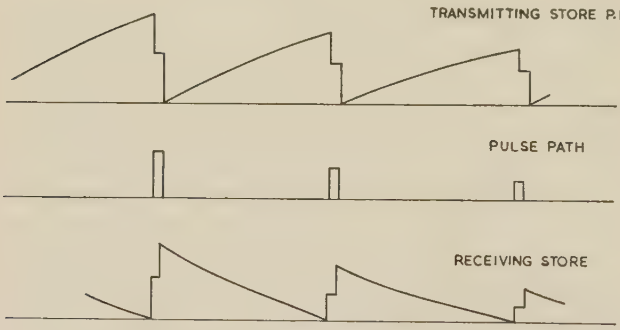


Fig. 5.—Charge, discharge and transfer waveforms.

depends on the complexity of the filter, approaching the theoretical limit more closely as the number of elements is increased (Fig. 4). Loss occurs not only through imperfection of the storage network but also through dissipation in the electronic switch. Experimentally, the best result obtained so far is a loss of about 2 dB, substantially constant over a 4 kc/s band with 10 kc/s sampling.

The waveforms observed at various points are shown in Fig. 5, for the case where delay lines are used in the storage networks. It is clear that current is being drawn from the audio source and fed to the audio load almost continuously, whereas current in the pulse path is concentrated into a short period. Charge accumulates slowly in the transmitting store, is transferred completely to the receiving store by a pulse of current and is then released slowly into the load.

The waveforms obtained with tuned circuits instead of delay lines in the stores are generally similar, except that the current pulse is a half-cosine wave and the steps in the waveform are replaced by smooth transitions. This is advantageous in that i reduces the bandwidth occupied by the signal, enables the effec of shunt capacitance on the common line to be reduced (or eve annulled, with certain combinations of circuit parameters) anc of course, uses fewer components. It has the drawback tha larger peak power is required to transmit a given mean signa power, and hence that the electronic switch must have a highe power rating.

(2.3) Reciprocity

It will be shown that the optimum storage networks are th same for transmitting and receiving. It is then clear from Fig. that the transmitting modulator and receiving demodulator ar exactly alike: the one device will perform either functior Following carrier system terminology, it can be called a 'modem

The significant properties of this pulse modem are that it ha a low transmission loss and is completely reciprocal. A p.a.m link can be made with no amplifiers and only one set of apparatu to work for both directions of transmission. Reciprocity is i principle distinct from efficiency, but is not much use without it although reciprocal amplifiers are known, their gain is severel limited by stability considerations.

It is therefore possible to connect telephone lines directly t the audio sides of a number of pulse modems whose pulse side are connected to a common line, and to set up temporary c permanent communication between any pair by synchronizin their switching pulses. This has obvious applications to mult plex transmission and switching. All conventional pulse modu lators or demodulators are irreversible, so that to obtain 2-wa communication either two independent paths and termin equipments are needed, as is usual in time-division multiple (t.d.m.) transmission systems, or two time positions must b provided for each channel, as has been proposed for t.d.n switching. Either course is comparatively lavish with apparatu and the second may be technically difficult if many channels a required.

Reciprocity has a further advantage in telephone switchin circuits. It is useful to be able to send dialling signals throug the speech path, to avoid the necessity for a separate path, which with electronic switching, is expensive to provide. This is mo readily done by conveying dialling pulses as modulation on voice-frequency tone. In normal multiplex paths this woul require an oscillator at the originating point; in the reciproca path, however, it is possible to use the v.f. equivalent of a centra battery. The tone can be supplied from a central point asso ciated with a detector and register. It passes through a modem in the central unit, through the pulse path, and through the sub scriber's modem to line. A mismatch, such as a short-circui reflects tone back into the modem; because this is reciprocal, th reflected tone travels back to the central unit, where it may b detected. Therefore a short-circuit (or, more generally, a impedance change) can be propagated through the entire speec path and detected at the tone source.[14]

(3) CHARGE INTERCHANGE IN STORAGE NETWORKS

(3.1) Delay Lines

The first mention of storage reactance in Section 2.1 wa deliberately couched in broad terms, since a wide variety o reactance networks can be pressed into service. Those teste experimentally have ranged from a single capacitance or indu tance to an assembly of 32 elements. However, the more usef arrangements all have a common feature. The function of th store is to absorb a charge slowly and discharge it rapidly, an although this can be accomplished by a single reactance connecte to alternative paths in conjunction with which it has differe

me-constants, the best results are obtained with a network which capable of discharging completely in a finite time.

A section of lossless transmission line, or a good lumped pproximation to such a line, has this property. If a charged ne section be connected to a resistive load matching its charac-ristic impedance, the terminal potential difference immediately ops to half the original open-circuit e.m.f.: this step of potential propagated along the line, reflected at the far end, and after period of twice the transmission delay returns to the near end, ducing the terminal p.d. to zero. The line is then completely scharged, and all the stored energy has been dissipated in the sistance load, at a constant rate for a short time. The process illustrated in Fig. 6.

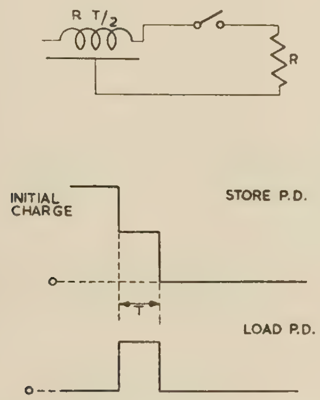


Fig. 6.—Discharge of delay line.

The delay line is the only network which discharges completely nd permanently in this manner. A single tuned circuit, and rtain other networks with an oscillatory discharge, pass rough an instantaneous condition of zero stored charge (not so zero stored energy, as with the line section), and provided at the switch through which the discharge takes place is closed r exactly half the period of oscillation, these can be used in a milar fashion.

In a reciprocal and symmetrical arrangement, such as Fig. 2, vo storage networks are connected together directly, and each rovides a matching termination to the other. If they are delay nes, of which one is initially charged, then a discharge takes lace as described above, for a period T, where the delay time $T/2$. No energy has been dissipated, and the entire charge has een transferred to the other line (Fig. 7). If the switch continued to conduct, the charge would oscillate between the two stores, but if it is opened after a time T the charge remains in the second store. During the connection, the terminal p.d. of both stores is half the initial e.m.f., but at the moment that the switch opens reflections raise the second store to the full potential and reduce the first one to zero.

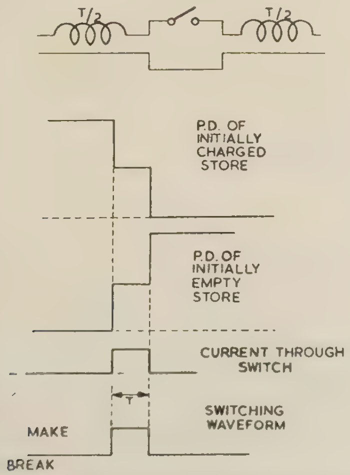


Fig. 7.—Interchange of charges of delay lines.

(3.2) Tuned Circuits

A pair of tuned circuits and their corresponding waveforms are shown in Fig. 8. The resonant frequency of the circuit is

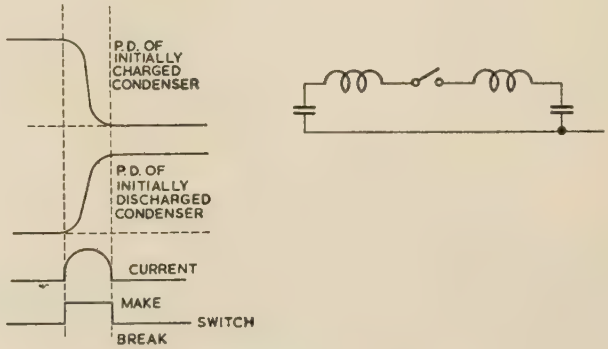


Fig. 8.—Interchange of charges of tuned circuits.

such that it executes one half-cycle of oscillation in the pulse period T, i.e.

$$\sqrt{(LC)} = \frac{T}{\pi} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The current flowing when the switch is closed is a half-sine wave, while the voltage across the storage capacitances are half-cosine waves in antiphase with each other and in quadrature with the current. If the peak voltage across either store is unity, the peak current is

$$I = \frac{1}{2}\sqrt{\frac{C}{L}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

As with a pair of line sections, the exchange of charge is complete, provided that there is no capacitance across the common line and that the switches are closed for precisely the period T. However, because the current rises gradually from zero at the start, and falls gradually to zero at the end of a pulse, imprecision of timing causes smaller errors than with delay-line stores. Also, the energy of the half-sine-wave current pulse is mainly at the lower end of the frequency spectrum, reducing the likelihood of induction between cables or components.

(3.3) Tuned Circuits and Line Capacitance

Thus there are two types of store which, when connected by electronic switching for a suitable short period, interchange their charges. If the pulse path has appreciable capacitance, the interchange is not complete, causing transmission loss, and a charge is left on the line after each pulse, causing crosstalk in a multiplex system. With tuned-circuit stores, the effect of line capacitance may, under certain conditions, be annulled.*

To find the waveforms in the presence of capacitance on the common line, we study the network of Fig. 9. The components L and C are the storage tuned circuits, while C′ simulates line capacitance. To simulate the closure of switches with the left-hand condenser C charged to unit potential, we assume an impulsive current of moment C, which instantaneously charges

* This method was suggested by R. B. Herman.[15]

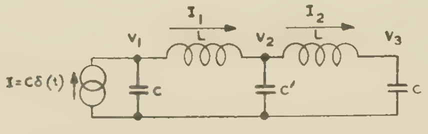


Fig. 9.—Equivalent network of stores and line capacitance.

the condenser and then leaves the network to execute its natural oscillations. It may be shown that

$$\left.\begin{aligned} v_1 &= \frac{k}{2k+1}U(t) + \tfrac{1}{2}\cos\alpha t + \frac{1}{2(2k+1)}\cos\beta t \\ v_3 &= \frac{k}{2k+1}U(t) - \tfrac{1}{2}\cos\alpha t + \frac{1}{2(2k+1)}\cos\beta t \end{aligned}\right\} \quad (3)$$

where

$$\left.\begin{aligned} k &= C/C' \\ \alpha &= 1/\sqrt{(LC)} \\ \beta &= \alpha\sqrt{(2k+1)} \end{aligned}\right\} \quad . \; . \; . \; . \; . \; . \quad (4)$$

and $U(t)$ is a unit step function.

If simultaneously $\cos\alpha t = -1$ and $\cos\beta t = 1$, the charges on the condensers C have been completely interchanged, since $v_1 = 0$ and $v_3 = 1$. Also, since the total charge on the two storage condensers is the same as at the beginning, there can be no charge on the line capacitance C' at this moment. The coincidence occurs when n cycles of $\cos\beta t$ occupy the same time as $m - \frac{1}{2}$ cycles of $\cos\alpha t$, namely when

$$\sqrt{(2k+1)} = \frac{\beta}{\alpha} = \frac{n}{m - \frac{1}{2}} \quad . \; . \; . \; . \quad (5)$$

where n and m are any positive integers. It follows that complete transfer can be effected, despite the presence of line capacitance, for an infinite number of capacitance ratios, so long as the resulting inductance values and waveforms are acceptable.

The case of practical significance is $m = n = 1$ with $k = 3/2$. The waveforms are then based on the half-sine wave, modified by some second harmonic, and are plotted in Fig. 10. Not only

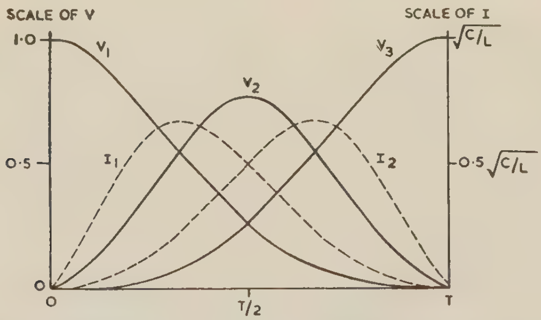


Fig. 10.—Interchange waveforms in presence of line capacitance.

are the transfer waveforms v_2, I_1 and I_2 zero at the beginning and end of the pulse; the derivative of v_2 at both bounds, and of the currents at one bound, are also zero: so that the high-frequency content of these waveforms is less, and the effect of mistuning less, than in either of the previously mentioned cases.

The component values are easily calculated from

$$\left.\begin{aligned} \sqrt{(LC)} &= \frac{T}{\pi} \\ \frac{C}{C'} &= \frac{3}{2} \end{aligned}\right\} \quad . \; . \; . \; . \; . \; . \quad (6)$$

if C' is given. If C' can be varied to some extent, the design would use this flexibility to obtain a desired impedance leve decided either on the basis of the pulse path impedance $\sqrt{(L/}$ or to suit the voltage and current limits of some particul electronic switch.

(3.4) The Approach to Impulsive Transfer

Any practical storage network for a pulse modem will conta some group of elements, either a delay line or a tuned circu which can discharge (or charge) completely within the short tim occupied by the pulse. The other half of the process, name charging (or discharging), occupies the relatively long perio between pulses, so that the remaining part of the network h time-constants of this order and its response during the pul period can be neglected. We shall find it convenient in develo ing the theory to treat the rapid charge or discharge as instan taneous, and to compute the effect of the pulsed connection o the low-frequency parts of the network by assuming it to b due to a train of ideal impulses of current. The approximatio is justified by the proportions of the waveforms observed, an by the correspondence between deductions from the theory an experimental results.

It may be remarked that storage networks inverse to tho shown here can be used, e.g. an open-circuit line discharge through a 'make' switch is equivalent to a short-circuit li discharged through a 'break' switch.

(4) THE RESPONSE OF NETWORKS TO MODULATED PULSE TRAINS

Some general concepts will be developed and applied to th analysis of various pulse communication systems.

(4.1) Impulsive Response and Steady-State Impedance

It is well known that an impedance may be specified by steady-state impedance function of the form

$$Z(p) = \frac{P(p)}{Q(p)} = \frac{kP(p)}{(p-p_1)(p-p_2)\ldots(p-p_r)} \quad . \quad (\,$$

or by the impulse response, namely the voltage waveform educe by an impulsive current of infinitesimal duration and unit mome applied at $t = 0$, related to $Z(p)$ by the inverse Laplace transfor

$$A(t) = \frac{1}{2\pi j}\int_{\gamma-j\infty}^{\gamma+j\infty} Z(p)\varepsilon^{pt}dp \quad . \; . \; . \; . \quad (\,$$

When the poles $p = p_r$ are all distinct, expansion of $Z(p)$ in partial fractions and application of eqn. (8) to each term giv

$$A(t) = \sum_r \frac{P(p_r)}{Q'(p_r)}\varepsilon^{p_r t} \quad . \; . \; . \; . \; . \quad (\,$$

where $Q' = dQ/dp$.

If an impulse of moment a_1 be applied at time t_1, the volta across the network at time t is

$$v(t) = a_1 A(t - t_1) \quad . \; . \; . \; . \; . \; . \quad (1$$

which, for a stable network, decays as $t \to \infty$.

(4.2) Impulse Trains

The response to a train of impulses is the sum of the respons to each individual impulse: if an impulse of moment a_n occu at time nt_1 where $(n \leqslant m)$, the voltage at time $(mt_1 + t)$ is

$$v(mt_1 + t) = \sum_n a_n A(t + mt_1 - nt_1) \quad . \; . \quad (1$$

and, if the a_n fall within finite bounds, this also decays as $t \to$ o

Now let the impulse train be modulated, with a sinusoidal envelope of frequency $\omega/2\pi$. This is most conveniently expressed as a complex exponential, so that

$$a_n = \varepsilon^{j\omega n t_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Assume that the response to the impulse train has reached a steady state. This implies that the impulse train has continued for an infinitely long time, so that

$$v(mt_1 + t) = \sum_{n=m}^{-\infty} \varepsilon^{j\omega n t_1} A(t + mt_1 - nt_1)$$

$$= \sum_{n=0}^{\infty} \varepsilon^{j\omega(m-n)t_1} A(t + nt_1)$$

$$= a_m \sum_{n=0}^{\infty} \varepsilon^{-j\omega n t_1} A(t + nt_1) \quad . \quad . \quad (13)$$

This equation may be interpreted in two ways. First, considering m as a parameter and t as a continuous variable, it gives the waveform in the period $mt_1 \leqslant t < (m+1)t_1$. The shape of the waveform is the same in each period, but its amplitude is proportional to that of the immediately preceding impulse. If the impulse train is unmodulated, $\omega = 0$ and $a_m = 1$ for all m; the waveform is then strictly periodic.

(4.3) Pulse-Sequence Impedance

Secondly, we may consider t as a parameter and m as a discrete variable. The expression then gives the envelope of the voltages obtained by sampling the waveform at corresponding parts of successive cycles. A particular case of great interest is obtained by sampling at mt_1, or, to be more precise, immediately after mt_1, since the effect of the pulse at mt_1 is included. If the amplitude of the sample is b_m, then

$$b_m = a_m \sum_{n=0}^{\infty} \varepsilon^{-j\omega n t_1} A(nt_1)$$

$$= a_m G(j\omega, t_1) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

defining a function $G(j\omega, t_1)$, which will be called the pulse-sequence impedance of the network.

Inserting the values of a_m from eqn. (12),

$$\text{Envelope of } a_m = \text{Modulating function} = \varepsilon^{j\omega t} \quad . \quad (15)$$

and

$$\text{Envelope of } b_m = \varepsilon^{j\omega t} G(j\omega) \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where the pulse-sequence impedance is written as a function of ω only to indicate that in all practical cases t_1 is fixed but ω is variable. The envelope* of the samples b_m is therefore a sinusoid of the same frequency as the modulating function; its relative amplitude and phase are given by the modulus and argument of $G(j\omega)$. The function exists for any stable network, since the series (14) converges if $A(t)$ decays exponentially as $t \to \infty$.

If the modulating waveform is not a sinusoid but is an arbitrary signal contained in the frequency range 0–$1/2t_1$, then, by superposition, the envelope of the samples (in the same frequency range) is derived from it by weighting its frequency spectrum with the function $G(j\omega)$, which can be treated like the frequency function of a normal filter. Ambiguity arises if higher modulating frequencies are applied, as it does in any sampling system.

* To be more rigorous, this is one of many envelopes, since a sinusoid of frequency $(r \pm \omega/2\pi)$ may be fitted to such a sequence of samples for all integral r.

Substituting $A(t)$ from eqn. (9) into eqn. (14),

$$G(j\omega) = \sum_{n=0}^{\infty} \varepsilon^{-j\omega n t_1} \sum_{r} \frac{P(p_r)}{Q'(p_r)} \varepsilon^{p_r n t_1}$$

$$= \sum_{r} \sum_{n=0}^{\infty} \frac{P(p_r)}{Q'(p_r)} \varepsilon^{nt_1(p_r - j\omega)}$$

$$= \sum_{r} \frac{P(p_r)}{Q'(p_r)} \frac{1}{1 - \varepsilon^{p_r t_1 - j\omega t_1}} \quad . \quad . \quad . \quad (17)$$

Another function which will be used is obtained by sampling the waveform immediately before the impulses are applied. This is

$$G_1(j\omega) = \sum_{n=1}^{\infty} \varepsilon^{-j\omega n t_1} A(nt_1)$$

$$= G(j\omega) - A(0)$$

$$= \sum_{r} \frac{P(p_r)}{Q'(p_r)} \frac{\varepsilon^{p_r t_1 - j\omega t_1}}{1 - \varepsilon^{p_r t_1 - j\omega t_1}} \quad . \quad . \quad . \quad (18)$$

As an example, the steady-state impedance of a capacitance C and resistance R in parallel is

$$Z(p) = \frac{1}{C} \frac{1}{p - p_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

where $p_1 = -1/(CR)$. Then

$$\left.\begin{aligned} G(j\omega) &= \frac{1}{C} \frac{1}{1 - \varepsilon^{p_1 t_1 - j\omega t_1}} \\ G(jy) &= \frac{1}{C} \frac{1}{1 - \varepsilon^{-\alpha - jy}} \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (20)$$

using a dimensionless parameter $\alpha = t_1/(CR)$ and a dimensionless frequency variable $y = \omega t_1$. Also

$$G_1(jy) = \frac{1}{C} \frac{\varepsilon^{-\alpha - jy}}{1 - \varepsilon^{-\alpha - jy}} \quad . \quad . \quad . \quad . \quad (21)$$

Both functions are periodic in ω with period $2\pi/t_1$. The magnitude of $G(jy)$ is given by

$$|G(jy)|^2 = G(jy)G(-jy) = \frac{1}{C^2} \frac{1}{1 - 2\varepsilon^{-\alpha}\cos y + \varepsilon^{-2\alpha}} \quad (22)$$

and oscillates between $1/C(1 \pm \varepsilon^{-\alpha})$. The complexity of the expressions increases rapidly with the number of elements in the network.

(4.4) Inversion of Pulse-Sequence Impedance

So far in this treatment of pulse-sequence response we have, as in the direct Fourier and Laplace integrals, transformed from a time to a frequency function with a given network. A practical problem might be presented the other way round: given some desired pulse-sequence frequency function, to find a suitable network. This process requires as its first step an inverse transformation.

One manifestation of the inherent bandwidth limit of sampling systems is that any pulse-sequence impedance is periodic in ω, with period $2\pi/t_1$. This is clear from the general form of $G(j\omega)$ in eqn. (17). It implies that the function can be expanded in a Fourier series of the form

$$G(j\omega) = \sum_{n} c_n \varepsilon^{jn\omega t_1} \quad . \quad . \quad . \quad . \quad (23)$$

But we have already used an expression of this form, namely eqn. (14); comparison shows that the coefficients c_n are, in fact,

the values of the impulse response at the sampling times $A(nt_1)$. It follows that

$$A(nt_1) = \frac{t_1}{2\pi}\int_0^{2\pi/t_1} G(j\omega)\varepsilon^{j\omega n t_1}d\omega \quad . \quad . \quad . \quad (24)$$

which is the desired inverse transformation. As an example, we take the pulse-sequence impedance of eqn. (20):

$$G(j\omega) = \frac{1}{C}\frac{1}{1-\varepsilon^{p_1t_1-j\omega t_1}} = \frac{1}{C}\sum_{m=0}^{\infty}\varepsilon^{-p_1mt_1-j\omega mt_1} \quad . \quad (25)$$

Then
$$A(nt_1) = \frac{t_1}{2\pi C}\sum_{m=0}^{\infty}\varepsilon^{-p_1mt_1}\int_0^{2\pi/t_1}\varepsilon^{j\omega(n-m)t_1}d\omega \quad . \quad (26)$$

The integral vanishes when $n \neq m$; the remaining integral is $2\pi/t_1$, so that

$$A(nt_1) = \frac{1}{C}\varepsilon^{-p_1nt_1} \quad . \quad . \quad . \quad . \quad . \quad (27)$$

This information does not specify the network or its impulse response uniquely. It might be thought at first that we could deduce that

$$A(t) = \frac{1}{C}\varepsilon^{-p_1t} \quad . \quad . \quad . \quad . \quad . \quad (28)$$

which is the impulse response of the RC network from which the calculation started. In fact, the impulse response could contain any other terms whatsoever which were identically zero at $t = nt_1$; the possible variety of such terms suggests that application of the inversion theorem to network synthesis may not be entirely straightforward.

(4.5) Historical

The foregoing technique, which was devised for dealing with the problems in pulse communication to be described, is closely related to methods which have been used in other fields. Pulse-sequence functions were first used in the analysis of servo systems with intermittent data,[1–4] and, more recently, somewhat similar methods have been used to obtain approximate solutions of continuous problems by computing at a finite number of sampling points only.[5, 6] MacColl, in the earliest published work on the subject,[1] attributes the original use of steady-state sampling theory to G. R. Stibitz and C. E. Shannon, at some time before 1944. Hurewicz[2] describes a pulse transfer function equivalent to $G_1(j\omega)$ and gives an extensive account of its applications in servo systems. Barker[4] gives the same name to a function equivalent to $G(j\omega)$ in our analysis, and appends a table giving $G(j\omega)$ pertaining to various $Z(p)$. Both writers use the variable $z = \varepsilon^{-j\omega t_1}$, which is more convenient for extensive manipulations of pulse-sequence functions. For use in combination with ordinary frequency functions, as in the present work, it is necessary to revert at some point to the direct frequency variable.

More recently, several papers have been published on what is now known as the 'z-transform', mostly in connection with servo systems.

(5) ANALYSIS OF TRANSMISSION CHANNELS

(5.1) The One-Way Channel

Fig. 11 shows the elements of a link containing a unilateral repeater which isolates the two modems and provides each with a matching termination. This simulates conditions in a link which is necessarily one-way, as in a radio system or a long line with repeaters. The link works between a source and a load of impedance R_1 over a transmission medium of impedance R_2, using pulses of length t_2 with a repetition period t_1. The storage

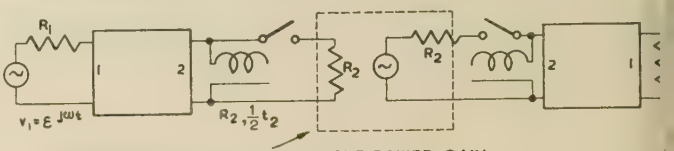


Fig. 11.—One-way channel.

network consists of a line of impedance R_2 and length $\frac{1}{2}t$ together with an arbitrary network represented by a box.

In all practical cases, the pulse length t_2 will be much smalle than the sampling period t_1, and the impedance R_2 of the pulse forming line will be much smaller than the signal source impe dance R_1 (i.e. $R_1/R_2 \simeq t_1/t_2$). Let us further assume that th impedances and time-constants of the network in the box show between the resistor R_1 and the line are of the order R_1 and t respectively, or at least that they are much greater than R_2 and t Then, to the same order of approximation, the delay line at th transmitter discharges completely into the load R_2 when th switch is closed. This discharge occurs so rapidly, compare with the recharging, that it can be taken as instantaneous; th waveform across the delay line is then the same as that acros an equivalent capacitor discharged by ideal impulses of curren Similarly, the receiver delay line can be taken as equivalent to capacitor which becomes instantaneously charged by impulsiv current.

The practical circuit can be replaced by the equivalent c

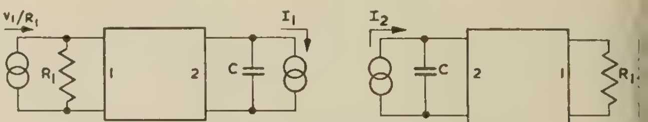


Fig. 12.—Equivalent of one-way channel.

Fig. 12. The currents I_1 and I_2 are modulated sequences c impulses

$$I_r = C\sum_n \varepsilon^{j\omega t}F_r(j\omega)\delta(t - nt_1) \quad . \quad . \quad . \quad (29)$$

where the functions $F_r(j\omega)$ are yet to be determined.

The effect of current I_1 acting alone can be found from th theory of Section 4. If the pulse-sequence impedance of th network of Fig. 13, viewed at terminals AA′, is $G(j\omega)$, th

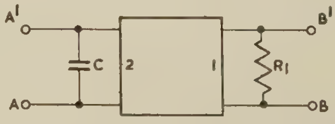


Fig. 13.—Section of equivalent network.

envelope of the potentials across C immediately after th impulses is $-C\varepsilon^{j\omega t}F_1(j\omega)G(j\omega)$. The envelope of potentials a the same point at the same times due to the modulating signa is $\varepsilon^{j\omega t}H_1(j\omega)/R_1$, where $H_1(j\omega)$ is the transfer impedance fror terminals BB′ to AA′ in Fig. 13. But, since the delay lin discharges fully, the potentials at these times are all zerc Equating the sum of the partial potentials to zero, we obtain

$$F_1(j\omega) = \frac{1}{R_1C}\frac{H_1(j\omega)}{G(j\omega)} \quad . \quad . \quad . \quad . \quad (30)$$

The envelope of the potentials on the transmitter delay lin immediately before the pulses is $\varepsilon^{j\omega t}F_1(j\omega)$. The transmitte pulses are half the amplitude (Fig. 14), and the repeater wit unity available power gain provides a source e.m.f. consistin

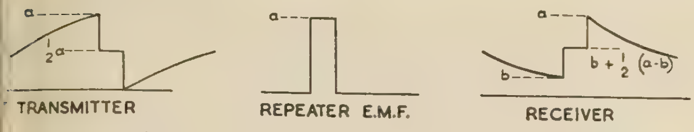


Fig. 14.—Waveforms in one-way channel.

of pulses with the full amplitude. The receiver delay line, whatever its residual charge, is charged up exactly to the full amplitude. Since the pulse sequence I_2 produces a sequence of potentials whose envelope is $\varepsilon^{j\omega t}F_1(j\omega)$, it follows that

$$F_2(j\omega) = \frac{1}{C}\frac{F_1(j\omega)}{G(j\omega)} \quad . \quad . \quad . \quad . \quad . \quad (31)$$

The component of current at frequency ω in I_2 is $CF_2(j\omega)/t_1$; it is this that produces an output voltage v_2 at signal frequency, so that

$$v_2 = \varepsilon^{j\omega t}\frac{C}{t_1}F_2(j\omega)H_2(j\omega) \quad . \quad . \quad . \quad . \quad (32)$$

where $H_2(j\omega)$ is the transfer impedance from terminals AA′ to BB′ in Fig. 14.

The voltage ratio produced by insertion of the whole system is $T(j\omega) = 2v_2/v_1$. Combining eqns. (30), (31), (32), and assuming a reciprocal network with $H_1 = H_2 = H$,

$$T_1(j\omega) = \frac{2}{t_1R_1C}\left[\frac{H(j\omega)}{G(j\omega)}\right]^2 \quad . \quad . \quad . \quad . \quad (33)$$

As an example, we take the elementary but perfectly practical case in which there is no intervening network between the pulse-forming lines and the audio source or load, and the pulse-sequence impedance $G(j\omega)$ is as in eqns. (20) and (22). The function $|T_1(j\omega)|$ is given in the fourth column and first row of Table 1, and plotted in Fig. 15 for various values of $\alpha = t_1/(CR)$ against the ratio between modulating and sampling frequencies.

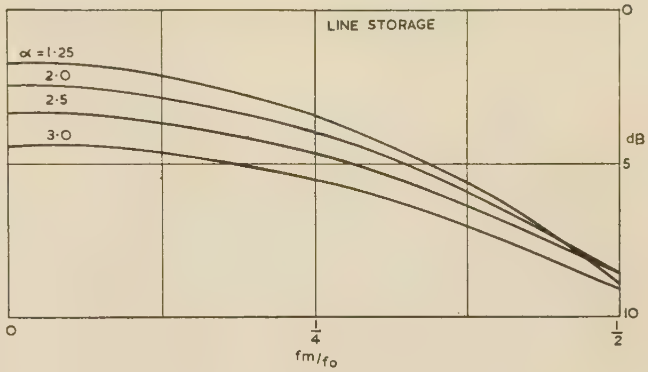


Fig. 15.—Transmission of a one-way channel with minimum storage network.

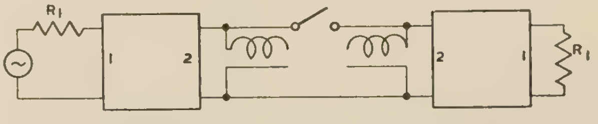


Fig. 16.—Two-way direct channel.

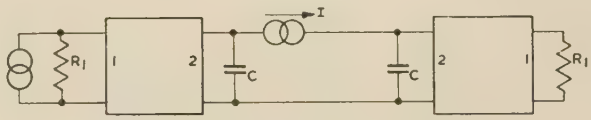


Fig. 17.—Equivalent of a two-way channel.

period t_2 which equals the length of pulse formed by one of the storage delay lines. While it is closed, the charge on one line flows into the other line; at the moment when the charge is about to return, the switch opens. On each operation of the switch the two stores exchange charges.

Table 1

Type	Schematic	General expression for transmission	Elementary modem: modulus of transmission	Transmission with ideal filter	
One-way	Fig. 12	$\frac{2}{t_1R_1C}\frac{H^2}{G^2}$	$\frac{4\alpha\varepsilon^{-\alpha}(\cosh\alpha - \cos y)}{\alpha^2 + y^2}$	1 0	$\lvert\omega\rvert < \pi/t_1$ $\lvert\omega\rvert > \pi/t_1$
Two-way direct . .	Fig. 16	$\frac{2}{t_1R_1C}\frac{H^2}{G^2 - G_1^2}$	$\frac{2\sqrt{2}\alpha(\cosh\alpha - \cos y)}{(\alpha^2 + y^2)(\cosh 2\alpha - \cos 2y)^{1/2}}$	1 0	$\lvert\omega\rvert < \pi/t_1$ $\lvert\omega\rvert > \pi/t_1$
Two-way with line of delay $mt_1/2$	Fig. 19	$\frac{2\varepsilon^{-\Upsilon - jm\omega t_1/2}}{t_1R_1C}\frac{H^2}{G^2 - \varepsilon^{-2\Upsilon - jm\omega t_1}G_1^2}$	$\frac{2\sqrt{2}\alpha(\cosh\alpha - \cos y)}{(\alpha^2 + y^2)[\cosh 2(\alpha + \gamma) - \cos(m+2)y]^{1/2}}$	$\varepsilon^{-\Upsilon - j\omega m t_1/2}$ 0	$\lvert\omega\rvert < \pi/t_1$ $\lvert\omega\rvert > \pi/t_1$
Two-way with intermediate storage	Fig. 20	$\frac{2\varepsilon^{-j\omega\tau}}{t_1R_1C}\frac{H^2}{G^2 - \varepsilon^{-j\omega t_1}G_1^2}$	$\frac{2\sqrt{2}\alpha(\cosh\alpha - \cos y)}{(\alpha^2 + y^2)(\cosh 2\alpha - \cos 3y)^{1/2}}$	$\varepsilon^{-j\omega\tau}$ 0	$\lvert\omega\rvert < \pi/t_1$ $\lvert\omega\rvert > \pi/t_1$
One-way into capacitive store	Fig. 22	$\frac{1}{CR_1}\frac{H}{G - \varepsilon^{-j\omega t_1}G_1}$	$\frac{\alpha(\cosh\alpha - \cos y)^{1/2}}{(\alpha^2 + y^2)^{1/2}(\cosh 2\alpha - \cos 2y)^{1/2}}$	1 0	$\lvert\omega\rvert < \pi/t_1$ $\lvert\omega\rvert > \pi/t_1$

(5.2) The Two-Way Direct Channel

Fig. 16 shows the elements of a system in which a number of modems are connected to a common busbar and communication is established between two of them by closing their switches in synchronism. According to the reasoning invoked to justify Fig. 12, this can be analysed by the equivalent circuit of Fig. 17. Since the two switches are directly in series and operate together, they can be replaced by one switch. This switch closes for a

The current I is a modulated sequence of impulses

$$I = C\sum_n \varepsilon^{j\omega t}F(j\omega)\delta(t - nt_1) \quad . \quad . \quad . \quad (34)$$

where the function $F(j\omega)$ has yet to be determined. The envelopes of the potentials across the lines immediately before and after the impulses may be found by using the pulse-sequence impedance functions. Because the lines interchange charges, the

potential across one of them immediately before the nth pulse equals the potential across the other immediately after the nth pulse, i.e.

$$\frac{H_1(j\omega)}{R_1} - CF(j\omega)G(j\omega) = CF(j\omega)G_1(j\omega) \quad . \quad . \quad (35)$$

which gives

$$F(j\omega) = \frac{1}{CR_1}\frac{H_1(j\omega)}{G(j\omega) + G_1(j\omega)} \quad . \quad . \quad . \quad (36)$$

The component of current at frequency ω in I is $CF(j\omega)/t_1$ which produces an output at signal frequency

$$v_2 = \varepsilon^{j\omega t}\frac{C}{t_1}F(j\omega)H_2(j\omega) \quad . \quad . \quad . \quad . \quad (37)$$

Combining eqns. (36) and (37) and assuming $H_1 = H_2 = H$, the overall voltage ratio is

$$T_2(j\omega) = \frac{2}{t_1R_1}\frac{[H(j\omega)]^2}{G(j\omega) + G_1(j\omega)} \quad . \quad . \quad . \quad (38)$$

To facilitate comparison with other systems we derive an alternative expression. For a network with shunt capacitance C, it can be shown from the definitions that

$$G(j\omega) - G_1(j\omega) = A(0) = \frac{1}{C} \quad . \quad . \quad . \quad (39)$$

and, combining this with eqn. (38),

$$T_2(j\omega) = \frac{2}{t_1R_1C}\frac{[H(j\omega)]^2}{[G(j\omega)]^2 - [G_1(j\omega)]^2} \quad . \quad (40)$$

The result for the elementary modem with no intervening network may be compared with that for one-way transmission from Fig. 18

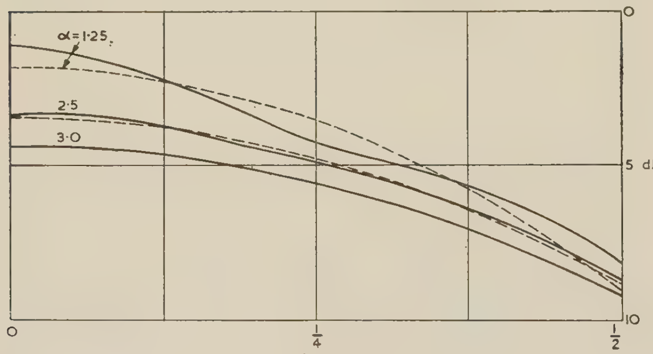


Fig. 18.—Transmission of a two-way channel with minimum storage network.

—— Bilateral. ---- Unilateral.

and Table 1. They differ by an oscillatory factor which covers two periods in the range $0 \leqslant y < 2$. The ripples introduced by this factor are small if $\cosh 2\alpha \gg 1$, and appear negligible if $\alpha > 2$.

(5.3) The Two-Way Delayed Channel

It is possible to operate reciprocally over a long line provided that the pulses transmitted from each terminal can be synchronized with those that it receives from the other. This requires that the transmission delay shall be a multiple of $t_1/2$. We shall consider the arrangement of Fig. 19, in which the two modems are connected by a line or network of delay $mt_1/2$, impedance R_2 (matching the storage lines) and loss γ nepers. The latter requires some explanation, since a practical line or network will not have a uniform loss at all frequencies in the

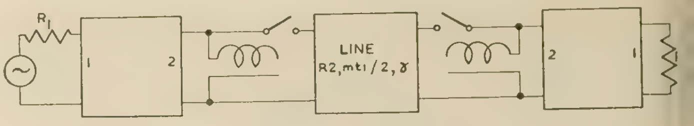


Fig. 19.—Two-way delayed channel.

signal spectrum. It is assumed that the line is equalized we enough to meet the crosstalk requirements of the system, s that most of the energy in a transmitted pulse is either dissipate in the line or received within the appropriate pulse period. Th loss is then to be considered as the ratio between the total charg received within the nominal pulse period and the total charg launched into the line as a transmitted pulse.

The equivalent circuit is the same as Fig. 12; I_1 and I_2 wi not have the same values as on the previous occasion when th was used (Section 5.1) but they are, of course, modulate sequences of impulses.

$$\left.\begin{aligned} I_1 &= C\sum_n \varepsilon^{j\omega t}F_1(j\omega)\delta(t - nt_1) \\ I_2 &= C\sum_n \varepsilon^{j\omega t}F_2(j\omega)\delta(t - nt_1 - mt_1/2) \end{aligned}\right\} \quad . \quad (4$$

At each closure of a switch, a storage line and the transmissio medium exchange charges (just as the two storage lines did wit direct connection). Taking as time reference a pulse from th transmitting end, the potential present on the transmitter sto immediately before this pulse will appear at the near end of th transmission line immediately after. It will be attenuated by factor $\varepsilon^{-\gamma}$ in the line, and arrives at the far end immediatel before time $mt_1/2$. The receiving switch closes at this time, an the potential is transferred to the receiver store. The potenti which had been in the receiver store immediately before $mt_1/2$ transferred to the line; after attenuation it arrives back at th transmitter at mt_1 and is transferred into the transmitter store.

Applying this reasoning to the envelopes of potentials foun by using pulse-sequence impedances, we can write

$$CG(j\omega)F_2(j\omega) = \varepsilon^{-\gamma - jm\omega t_1/2}\left[\frac{H_1(j\omega)}{R_1} - CG_1(j\omega)F_1(j\omega)\right]$$

$$\frac{H_1(j\omega)}{R_1} - CG(j\omega)F_1(j\omega) = \varepsilon^{-\gamma - jm\omega t_1/2}CG_1(j\omega)F_2(j\omega) \quad (4$$

which can be solved for $F_2(j\omega)$.

The component of received current at frequency ω $CF_2(j\omega)/t_1$: this is modified by a factor $H_2(j\omega)$ before reachi the output. So the overall transmission is (with reciproc networks)

$$T_3(j\omega) = \frac{2\varepsilon^{-\gamma - jm\omega t_1/2}}{CR_1t_1}\frac{[H(j\omega)]^2}{[G(j\omega)]^2 - \varepsilon^{-2\gamma - jm\omega t_1}[G_1(j\omega)]^2} \quad (4$$

in which $\varepsilon^{-jm\omega t_1/2}$ is a delay operator corresponding to th transmission time of the cable. It will be seen that if $\gamma = m =$ this expression reduces to eqn. (40) for the direct connectio For the elementary modem with no intervening network, Table shows that ripples of closer spacing than with direct connectic are introduced into the frequency characteristic, with amplitu which is negligible for $\alpha + \gamma > 2$.

(5.4) The Two-Way Channel with Intermediate Storage

In some multiple-stage switching systems it is necessary t convert a signal in the form of a modulated pulse train in certain time phase to a similar train in a different time phas This may be accomplished by recovery of the audio signal ar remodulation, or by the use of delay networks suitab

switched; an extension of the modem technique, however, provides a more elegant and economical solution.

The method uses two successive charge exchanges.* By closure of the electronic switch S1 (Fig. 20) at times nt_1 the

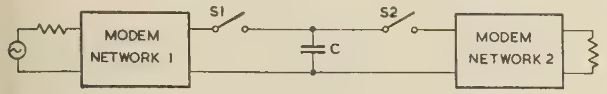


Fig. 20.—Two-way channel with intermediate storage.

modem network 1 exchanges charges with the storage condenser C. At later times $nt_1 + \tau$, by closure of S2, the condenser exchanges charges with modem network 2. The analysis of the successive interchanges resembles that of the delayed system treated in Section 5.3, and yields the transmission function from modem 1 to modem 2:

$$T_4(j\omega) = \frac{2}{CRt_1}\varepsilon^{-j\omega\tau}\frac{[H(j\omega)]^2}{[G(j\omega)]^2 - \varepsilon^{-j\omega t_1}[G_1(j\omega)]^2} \quad (44)$$

The reverse transmission is identical except that the delay operator $\varepsilon^{-j\omega\tau}$ is replaced by $\varepsilon^{-j\omega(t_1-\tau)}$, which for speech signals is not significant. For comparison with the other systems see Table 1.

(5.5) One-Way Transmission into a Capacitive Store

It is sometimes required that a potential proportional to an instantaneous signal value, obtained from a sample of short duration, shall be maintained during the interval between successive samples, so producing what has been called a 'box-car' waveform (Fig. 21). This is normally done by charging a

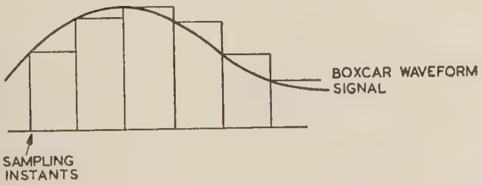


Fig. 21.—Box-car waveform.

capacitor as rapidly as possible to the sample potential and isolating it during the interval. The large current pulse required for rapid charging is economically generated by discharge of a modem-type network.

Let the signal be applied through a storage network, as shown in Fig. 22, to a sampling switch making connection to the capaci-

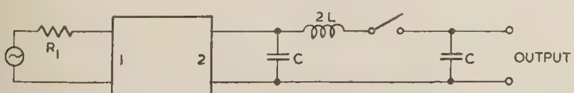


Fig. 22.—One-way transmission into capacitive store.

tor across which the output potential is desired. The storage network in the diagram employs a tuned circuit whose parameters are related to the pulse width by eqn. (1), and with a capacitance C equal to that of the holding capacitor. The waveforms occurring during sampling are those shown in Fig. 8.

The envelope of the sample amplitudes stored on the output capacitor must be a reasonable replica of the signal. A transfer function relating the two may be found by use of an equivalent circuit, Fig. 23, which is similar to Fig. 17 except that instead of a modem on the right-hand side there is simply a storage condenser. The current I is a sequence of impulses, with an envelope

* This method is due to W. Bezdel.[16]

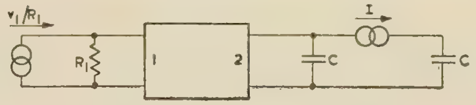


Fig. 23.—Equivalent of transmission into capacitive store.

function $F(j\omega)$ as in eqn. (34). By interchange of charges, the potential on the modem just before the nth pulse is transferred to the condenser, is held between the nth and $(n + 1)$th pulses, and reappears on the modem just after the $(n + 1)$th pulse, i.e.

$$\frac{H(j\omega)}{R_1} - CF(j\omega)G_1(j\omega) = \varepsilon^{-j\omega t_1}\left[\frac{H(j\omega)}{R_1} - CF(j\omega)G(j\omega)\right] \quad (45)$$

whence the desired transfer function is

$$T_5(j\omega) = CF(j\omega)G_1(j\omega) = \frac{1}{CR_1}\,\frac{H(j\omega)}{G(j\omega) - \varepsilon^{-j\omega t_1}G_1(j\omega)} \quad (46)$$

The expression for the elementary modem appears in the last row of Table 1.

(5.6) Comparison of the Five Systems

In the early part of this Section, the transmission properties of five channels employing storage in one way or another have been derived in terms of the pulse-sequence functions defined in Section 4. For reference and easy comparison, the general expressions are repeated in Table 1, together with the specific expressions for the elementary modem having no filter network but only a delay line or tuned circuit to provide a pulse discharge. There is a strong generic resemblance between the five expressions of each group, and, as will be discussed in the next Section, a network which is good for one system is, in fact, good for all.

(6) SYNTHESIS OF THE OPTIMUM STORAGE NETWORK

(6.1) The Ideal Limit

The final aim of analysis is synthesis: we desire not merely to know what a given system will do but to know what system will have a given performance. The quantity which we should like to specify in this case is the overall channel transmission of a pulse link, given in the third column of Table 1 for the five types system considered. Network synthesis at its best is something of a black art, and synthesis for a complicated mixture of steady-state and impulse properties might be thought a hopeless task. A general procedure for obtaining any given $T(j\omega)$ is probably in this category, but a result can be obtained for a specific ideal transmission function which represents a limit to the attainable performance, and to which approximations of various orders can be made.

The transmission modulus $|T_1(j\omega)|$ for the one-way system is the simplest starting-point. Let us further specify the storage networks to contain only lossless reactances. Since both theory and experiment so far point to some form of filter as the best thing to place in the box between the delay lines and the resistance R_1 (Figs. 11–13, 16, 17, 19, 22, 23) this is reasonable; and it allows the transfer impedance $|H(j\omega)|$ to be related to a driving point impedance, which simplifies the problem. Fig. 13 then consists of a lossless 2-port network terminated in a simple resistance R_1. If this is driven from end AA′, the power dissipated in R_1 must equal the power fed into AA′. Writing the real part of the input impedance $Z(j\omega)$ at AA′ as $R(\omega)$, then[7]

$$R_1R(\omega) = |H(j\omega)|^2 \quad (47)$$

Provided that this input impedance is a minimum-reactance function, knowledge of $R(\omega)$ determines $Z(j\omega)$ and hence $Z(p)$.

These may be calculated explicitly,[7] but our immediate purpose is to use this fact indirectly to derive the impulsive response $A(t)$ from $R(\omega)$ alone. The result may be quoted from Goldman:[8]

$$A(t) = \frac{2}{\pi}\int_0^{\infty} R(\omega)\cos\omega t d\omega$$

whence also

$$A(t) = \frac{1}{\pi}\int_{-\infty}^{+\infty} R(\omega)\varepsilon^{j\omega t}d\omega \quad . \quad . \quad . \quad . \quad (48)$$

the second form being justified by the fact that $R(\omega)$ and $\cos\omega t$ are even functions, while $\sin\omega t$ is an odd function.

Combining eqns. (33) and (47),

$$R(\omega) = \frac{t_1 C}{2}|T_1(j\omega)|\,|G(j\omega)|^2 \quad . \quad . \quad . \quad (49)$$

which may be inserted in eqn. (48) to obtain an expression for $A(t)$. We can also obtain discrete values of $A(t)$ in terms of $G(j\omega)$ from the inverse transform, eqn. (24), and these must be consistent with the continuous expression. This gives an infinite set of integral equations

$$\frac{2\pi}{t_1}A(nt_1) = \int_{-\pi/t_1}^{+\pi/t_1} G(j\omega)\varepsilon^{j\omega nt_1}d\omega = C\int_{-\infty}^{+\infty}|T_1(j\omega)|\,|G(j\omega)|^2\varepsilon^{j\omega nt_1}d\omega \quad . \quad . \quad . \quad . \quad (50)$$

relating $G(j\omega)$ to the specified transmission $|T_1(j\omega)|$. If these can be solved for $G(j\omega)$, then $A(nt_1)$ follows and the network synthesis is reduced to a known problem.

At this point we introduce the ideal transmission characteristic. A sampling system is fundamentally limited to a useful bandwidth of half the sampling rate, namely $1/2t_1$; we shall therefore postulate a constant modulus of transmission up to this frequency and zero beyond, i.e.

$$\begin{aligned} |T_1(j\omega)| &= k \qquad 0 < |\omega| < \pi/t_1 \\ &= 0 \qquad \pi/t_1 < |\omega| \quad . \quad . \quad . \quad . \quad (51) \end{aligned}$$

It is to be hoped that the system can be made lossless over this bandwidth, i.e. $k = 1$. Substituting condition (51) into eqn. (50),

$$\int_{-\pi/t_1}^{+\pi/t_1} G(j\omega)\varepsilon^{j\omega nt_1}d\omega = kC\int_{-\pi/t_1}^{+\pi/t_1}|G(j\omega)|^2\varepsilon^{j\omega nt_1}d\omega \quad . \quad (52)$$

for all integral n. Consider the physical meaning of this set of equations. The left-hand integral gives a coefficient in the Fourier expansion of $G(j\omega)$, which we know is periodic in ω with period $2\pi/t_1$. The right-hand integral gives a coefficient of the same order in the Fourier expansion of $|G(j\omega)|^2$. The equation states that corresponding coefficients in the two series are proportional. $G(j\omega)$ is therefore a function which, when squared, retains the same spectrum and therefore the same shape. It could be a constant or it could be any function alternating beween a constant and zero. Since $G(j\omega)$ is in the denominator of $T(j\omega)$, a zero value is inadmissible; it must therefore be a constant:

$$G(j\omega) = \frac{1}{kC} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$

Applying the inverse transform, eqn. (24), we find that

$$\left.\begin{aligned} A(nt_1) &= \frac{t_1}{2\pi kC}\int_0^{2\pi/t_1}\varepsilon^{j\omega nt_1}d\omega \\ A(0) &= \frac{1}{kC} \\ A(mt_1) &= 0 \qquad\qquad m \neq 0 \end{aligned}\right\} \quad . \quad . \quad . \quad (54)$$

The instantaneous potential assumed by a network with shunt capacitance C when a unit impulse is applied cannot exceed $1/C$; it attains this if all the impulsive current flows into the terminal capacitance, as it should in a low-pass ladder network. The constant k, being the transmission ratio of a passive system, cannot exceed unity: it must therefore be unity.

Also, the impulse response exists at $t = 0$ but vanishes at all other sampling instants. It is not possible for the impulse response of a network with shunt capacitance to be identically zero for a finite period of time, so $A(t)$ must oscillate with period $2t_1$; this is characteristic of a low-pass filter with cut-off frequency $1/2t_1$. Such a filter is in fact specified, since with constant $G(j\omega)$ the required transmission $T(j\omega)$ must be shaped by the network transfer impedance. From eqns. (33) and (53),

$$\begin{aligned} |H(j\omega)|^2 &= \frac{t_1R_1}{2C} \qquad 0 < \omega < \pi/t_1 \\ &= 0 \qquad\qquad \pi/t_1 < |\omega| \quad . \quad . \quad . \quad (55) \end{aligned}$$

This transfer impedance determines the real component of driving-point impedance [eqn. (47)] which in turn determines the driving-point impulse response [eqn. (48)]. This is

$$A(t) = \frac{t_1}{\pi C}\int_0^{\pi/t_1}\cos\omega t d\omega = \frac{1}{C}\,\frac{\sin(\pi t/t_1)}{\pi t/t_1} \quad . \quad . \quad (56)$$

a decreasing oscillatory function which accords with the conditions imposed above.

At low frequencies, the transfer impedance of the low-pass filter necessarily tends towards R_1; so its constant value over the pass band must equal R_1. This implies that

$$R(\omega) = |H(j\omega)| = R_1 = t_1/2C \quad . \quad . \quad . \quad (57)$$

It might be thought necessary to verify that such a value of the real component can be sustained over a bandwidth $1/2t_1$ in the presence of a shunt capacitance C. In fact, this has been done implicitly by the calculation of impulse responses from the modified Fourier transform (48), since this takes into account the essential relation between real and imaginary parts of an impedance. To check the point explicitly, we may use Bode's resistance-integral theorem[9], which states that

$$\int_0^{\infty} R(\omega)d\omega = \frac{\pi}{2C} \quad . \quad . \quad . \quad . \quad . \quad (58)$$

and which is satisfied by the resistance, capacitance and bandwidth specified.

The network characteristics which have been calculated to obtain ideal transmission $|T_1(j\omega)|$ over a one-way system also result in ideal transmission $|T_2(j\omega)|$ over a 2-way system with direct connection; transmission $|T_3(j\omega)|$ for the delayed 2-way system which falls short of the ideal only by the inescapable delay and attenuation of the transmission line; transmission $|T_4(j\omega)|$ via an intermediate store with only the delay between input and output pulse trains; and an ideal transfer ratio $|T_5(j\omega)|$ into a holding capacitor. (See the last column of Table 1.) In the first four cases, since the full available power is drawn from the modulation source, that source is perfectly matched; by symmetry, the signal load impedance at the receiver is perfectly matched also.

We have found a specification for a network to give the best possible transmission, namely uniform lossless transmission over a bandwidth equal to half the sampling rate.[17] The combination of the line capacitance C with the unknown network in the box in Fig. 13 must constitute an ideal low-pass filter, the real part of whose input impedance is constant over the pass band and

zero elsewhere. This, like any other characteristic with a discontinuity, or with infinite attenuation over a continuous range of frequencies, is not strictly realizable. It is an ideal to which practical approximations can be made. We may hope that a good approximation to the ideal filter will give a good approximation to the ideal transmission: this is a matter requiring further theoretical and experimental study. Theoretical possibilities are discussed in the remainder of this Section, and experimental results are given in a companion paper.[13]

(6.2) Properties of Useful Approximations

A lossless reactance network which approximates the ideal transfer impedance of eqn. (55) will also approximate the ideal driving-point impedance and impulse response. It is probable that a high-order approximation will serve the purpose regardless of its design basis, but we are normally more interested in the departures from the ideal resulting from the use of networks with only a few elements.

Since the exact computation of the transmission functions is very laborious in all but the simplest cases, it is advisable to take a first glance at the problem in a rough-and-ready manner. At low frequencies, the transfer impedance $H(j\omega) \simeq R_1$, and with any useful filter characteristics it will vary smoothly in the vicinity of this value over a large part of the pass band. The pulse-sequence impedance $G(j\omega)$ is a periodic function, the coefficients of whose Fourier series are $A(nt_1)$, the values of the impulse response at sampling times. If, as desired, these are small, and if the series converges rapidly, the first two or three terms should suffice to give some idea of the transmission function. The zero-order coefficient is always $1/C$, so the mean transmission level within the pass band will be approximately $2R_1C/t_1$ with either one-way or 2-way transmission. An ideal filter of bandwidth $1/2t_1$ and impedance R_1 has a terminal capacitance such that this is unity; a practical filter has a lower capacitance, which will lower the transmission level. We require a filter with as high a terminal capacitance as possible, and with the impulse response small at the first few sampling instants.

It will be recalled that the transmission characteristic expected is the product of three terms, a filter transfer function, a constant defining the mean level, and a periodic function with ripple amplitudes depending on the impulse response of the network. For one-way transmission the latter function is

$$\frac{C^2}{|G(j\omega)|^2} \simeq [(1 + A_1^2) + 2A_1(1 + A_2)\cos y + 2A_2 \cos 2y + 2A_3 \cos 3y + \ldots]^{-1} \quad . \quad (59)$$

and for 2-way transmission,

$$\frac{C}{G(j\omega) + G_1(j\omega)} \simeq [(1 + 4A_1^2) + 4A_1(1 + 2A_2)\cos y + 4A_2 \cos 2y + 4A_3 \cos 3y + \ldots]^{-1/2} \quad (60)$$

where A_m is the normalized impulse response $A(mt_1)/A(0)$ at the mth sampling instant, and terms beyond third order are neglected. The coefficients in either expansion are about twice those in the expansion of $G(j\omega)$.

(6.3) Maximally-Flat and Equal-Ripple Filters

We first examine some well-known ladder filters designed on an insertion-loss basis.[10, 11]

Two types of approximation to a filter flat over an angular pass band ω_c and cutting off sharply thereafter combine useful properties with mathematical simplicity. These are the maximally-flat, or Butterworth, characteristic and the equal-ripple, or Chebyshev, characteristic. The maximally-flat ladder net-

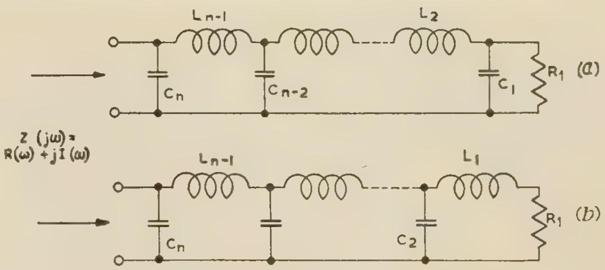


Fig. 24.—Low-pass ladder network.
(a) Odd number of elements. (b) Even number of elements.

work with n elements (Fig. 24) has a transfer modulus (hence real component of driving-point impedance)

$$R_n(\omega) = \frac{R_1}{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}} \quad . \quad . \quad . \quad . \quad (61)$$

The terminal capacitance is

$$C_n = \frac{n}{\omega_c R_1} \sin \frac{\pi}{2n} \quad . \quad . \quad . \quad . \quad (62)$$

so that $\omega_c R_1 C_n$ tends to $\pi/2$ for large n, thus approximating to our desired pass band, resistance and capacitance. For the equal-ripple ladder network with n elements,

$$R_n(\omega) = \frac{hR_1}{1 + k^2 \cos^2\left(n \cos^{-1} \frac{\omega}{\omega_c} \cosh \beta\right)} \quad . \quad (63)$$

where k is a parameter determining the ripple amplitude, β is a related parameter defined by $\sinh n\beta = 1/k$, and h is a factor equal to unity for odd values of n, and $1 + k^2$ for even values. The terminal capacitance is

$$C_n = \frac{n\sqrt{(k^2 + 1)}}{h\omega_c R_1} \sin \frac{\pi}{2n} \quad . \quad . \quad . \quad . \quad (64)$$

which differs from that of the maximally-flat filter by a factor $(1 + k^2)^{\pm 1/2}$.

From eqn. (62) the capacitance of a maximally-flat filter approaches the theoretical limit fairly closely for the third and higher orders; while from eqn. (64) the capacitances of the equal-ripple filters approach the same values as the ripple amplitude is reduced. It is not at all certain that the impulse response is satisfactory, since this is composed of damped oscillations at frequencies well below cut-off. We must therefore examine the impulse response of a ladder filter, considered as a 2-terminal impedance 'viewed' at the unloaded end. Although these networks are well known, little information about their impulsive properties has been published; and the numerical results which follow are obtained from the formulae quoted in the Appendix.

Mean levels and the first few coefficients A_m have been calculated for maximally flat networks of orders 2–5 and are given in Table 2.

Table 2

Number of elements n	Mean level $2R_1C/t_1$	Coefficients in expansion of $G(j\omega)$: $A_m = \frac{A(mt_1)}{A(0)}$			
		A_1	A_2	A_3	A_4
2	0·900	0·0206	−0·0145	0·0017	−0·0001
3	0·955	−0·0001	−0·0125	0·0060	−0·0017
4	0·974	−0·0058	−0·0078	0·0065	−0·0033
5	0·984	−0·0073	−0·0038	0·0072	−0·0049

It will be seen that convergence is rather slow; but all the terms are fairly small, which suggests that this type of approximation is sufficiently good to warrant a serious trial.

Similar coefficients have been calculated for equal-ripple filters of four elements with various amplitudes of ripple, and are given in Table 3. These present no very obvious advantage

Table 3

Percentage ripple	$b = \tanh \beta$	Coefficients in expansion of $G(j\omega) : A_m = \frac{A(mt)}{A(0)}$			
		A_1	A_2	A_3	A_4
0 (m.f.)	1	−0·0058	−0·0078	0·0065	−0·0033
0·167	0·75	−0·0129	−0·0023	0·0137	−0·0072
5	0·49	−0·0220	0·0100	−0·0300	−0·0080
10	0·42	−0·0208	−0·0167	−0·0217	−0·0103

over the maximally-flat filter in this respect: in fact, the coefficients generally increase with increase of ripple amplitude. However, the difference is not sufficient to exclude the equal-ripple filter, and the sharper cut-off obtainable with a given number of elements favours the use of this type if possible. Consequently both types of filter have been used in experimental work.

(6.4) Approximations: Other Considerations

The inversion of a pulse-sequence impedance does not specify a network or its impulse response uniquely. The inverse transform of a constant, eqn. (54), is no exception: it states that the impulse response shall be non-zero at $t = 0$ and zero at later sampling times $t = mt_1$, but it does not prescribe the behaviour at intermediate times. In consequence, there exists a class of filters having the desired property, of which the ideal filter with a rectangular frequency response is but one member. This class includes all low-pass filters whose real component of input impedance $R(\omega)$ (equal to the transfer modulus squared) has a cut-off shape which is skew-symmetric about a cut-off frequency ω_c equal to half the sampling rate. Such a curve can be thought of as the sum of a step at ω_c and an odd function of $(\omega - \omega_c)$, as illustrated in Fig. 25. The property is proved in the Appendix.

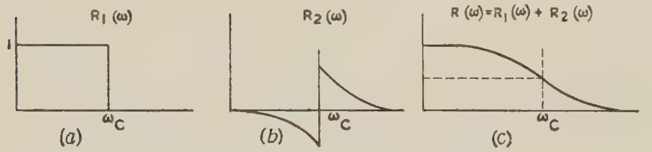


Fig. 25.—Skew-symmetric filter functions.
(a) Rectangular component. (b) Odd component. (c) Skew-symmetric function.

A transmission system employing this type of filter in its modems therefore has an overall transmission modulus determined entirely by the filter function, since if we substitute $G(j\omega) = 1/C$ and $|H(j\omega)|^2 = R_1 R(\omega)$ in either eqn. (33) (for one-way working) or eqn. (40) (for 2-way working) the result is

$$|T(j\omega)| = \frac{2C}{t_1} R(\omega) = \frac{R(\omega)}{R(0)} \quad . \quad . \quad . \quad (65)$$

The low-frequency transmission is unity, as with the square-cut-off filter, because for a given $R(0)$ and ω_c the resistance integral is the same for all curves with the specified type of symmetry. (Clearly the function $R_2(\omega)$ in Fig. 25 contributes nothing to this integral.)

Consequently, if a transmission function having this type of symmetry is desired (or can be tolerated), the synthesis problem is reduced to that of designing a passive network with the specified response shape. This, like the square cut-off is not strictly realizable, as can be seen at once from the fact that it has infinite attenuation over a range of (at least) $\omega \geqslant 2\omega_c$. Presumably the best that can be done by a filter of the configuration shown in Fig. 24 is an approximately skew-symmetric transition between zero loss at low frequencies and a large but finite attenuation at $2\omega_c$, with an asymptotic cut-off slope of $6n$ decibels per octave beyond. The corresponding impulse response has been computed (see Appendix); although not passing through zero at the sampling instants, it assumes values which are substantially lower than for the well-known filter types and which are negligibly small after the first few points. We conclude that the degree of approximation of a realizable filter can be good enough for practical purposes. The remaining problem, of synthesizing a filter with the desired cut-off, is by no means trivial, and in fact a preliminary attempt has failed to lead to a useful design: it is, however, a problem of known type, to which a solution should be obtainable if a sufficient effort is made.

It should be clearly understood that there is no inherent virtue in a skew-symmetrical cut-off: in fact, it could be argued that the resulting 6 dB attenuation at half the sampling rate is inadequate. The idea is introduced because it provides a possible approach to an otherwise rather intractable problem.

The most useful filter design would, of course, be one which caused the overall transmission to approximate the ideal in some specified manner. That this has not been accomplished is due to the complexity of the exact expression for overall transmission in any but the simplest cases. It is fortunate that filter networks of the known types do, in fact, yield a performance good enough to make the use of the modem entirely practical.

(7) REVIEW OF METHODS AND APPLICATIONS

The simplicity and efficiency of the pulse modem make it attractive for almost any application of pulse amplitude modulation, even where the reciprocity cannot also be utilized. It has therefore been considered for application to direct p.a.m. transmission on f.m. radio links, and to intermediary p.a.m. in systems using other time-division methods.[13]

However, these do not fully utilize its properties. The purpose of these concluding remarks is to outline the properties of the pulse modem and to stress the economy and elegance which it appears to offer in telephone switching applications.

The modem converts speech power to pulses, and a pulse train back to speech, very efficiently. In principle, the system can transmit signals falling within a bandwidth of half the sampling rate entirely without loss, and in both directions. Practical experiments have shown an overall loss of 2 dB, which it is hoped to reduce to about 1 dB when better switches are available, and a bandwidth of 4 kc/s with 10 kc/s sampling. The resulting telephone quality is subjectively pleasant.[13] Voice-frequency signals can be conveyed through a speech path very economically by utilizing the reciprocity of the system. With a particular modem design, the effect of a fairly large shunt capacitance on level and crosstalk can be rendered negligible, permitting the use of a large number of gates in parallel, perhaps 250 gates being practicable, a number unlikely to be exceeded in a 4-digit non-demodulating exchange. Level stability is good, since the system has an inherently low loss: there are no amplifiers, and with a certain modem design the effect of small errors of timing in the pulse path is negligible. In conjunction with suitable transistor circuits,[18] the modem provides a good electrical analogue of a rotary switch, although it should not be considered

as essentially associated with any particular device or control circuit.

It can be asserted with some confidence, therefore, that the efficient reciprocal circuit is a significant development in time-division technique with many possibilities in the field of telephone switching.*

(8) ACKNOWLEDGMENTS

The author is indebted to Messrs. H. Grayson, J. C. Price and R. B. Herman for discussion and suggestions concerning this work, and to Standard Telecommunication Laboratories Ltd. for permission to publish the paper.

(9) REFERENCES

(1) MacColl, L. A.: 'Fundamental Theory of Servomechanisms' (Van Nostrand, New York, 1945), Chapter 10.

(2) James, H. M., Nichols, N. B., and Phillips, R. S. (Ed.): 'Theory of Servomechanisms' (McGraw-Hill, New York, 1947), Chapter 5, Hurewicz, W., 'Filters and Servo Systems with Pulse Data'.

(3) Lawden, D. F.: 'A General Theory of Sampling Servo Systems', *Proceedings I.E.E.*, Monograph No. 4 M, July, 1951 (**98**, Part IV, p. 31).

(4) Barker, R. H.: 'The Pulse Transfer Function and its Application to Sampling Servo Systems', *ibid.*, Monograph No. 43 M, July, 1952 (**99**, Part IV, p. 302).

(5) Lewis, N. W.: 'Waveform Computation by the Time-Series Method', *ibid.*, Paper No. 1382 R, September, 1952 (**99**, Part III, p. 294).

(6) Truxal, J. G.: 'Numerical Analysis for Network Design', *Transactions of the Institute of Radio Engineers*, September, 1954, *CT*-**1**, p. 49.

(7) Guillemin, E. A.: 'Advances in Electronics, 3: Modern Methods of Network Synthesis' (Academic Press, New York, 1951), p. 275.

(8) Goldman, S.: 'Transformation Calculus and Electrical Transients' (Prentice-Hall, New York, 1949), p. 125.

(9) Bode, H. W.: 'Network Analysis and Feedback Amplifier Design' (Van Nostrand, New York, 1945), p. 280.

(10) Orchard, H. J.: 'Formulae for Ladder Filters', *Wireless Engineer*, January, 1953, **30**, p. 3.

(11) Belevitch, V.: 'Tchebycheff Filters and Amplifier Networks', *ibid.*, April, 1952, **29**, p. 106.

(12) Sunde, E. D.: 'Theoretical Fundamentals of Pulse Transmission', Bell System Monograph No. 2284.

(13) Price, J. C.: 'Efficiency and Reciprocity in Pulse Amplitude Modulation: Part 2—Testing and Application' (see page 463).

(14) Cattermole, K. W.: British Patent Application No. 38289: 1956.

(15) Cattermole, K. W., *et al.*: British Patent Application No. 17802: 1956.

(16) Wright, E. P. G., and Bezdel, W.: British Patent Application No. 17395: 1956.

(17) Cattermole, K. W., *et al.*: British Patent Application No. 17802: 1956.

(18) Cattermole, K. W.: 'Transistor Pulse Generators for Time-Division Multiplex' (see page 471).

(19) Haard, H. B., and Svala, C. G.: U.S. Patent No. 2718621: 1953.

* At a late stage of this work, which started early in 1953, it came to light that a circuit essentially like the 2-way direct arrangement discussed in Section 5.2 had been independently adopted in two other laboratories. The earliest invention appears to be that of Haard and Svala.[19] The Post Office Research Station exhibited another version at the Physical Society Exhibition, March, 1956. All parties appear to agree that the major application is to telephone switching. The other channels considered, the annulment of line capacitance and the whole of the detailed analysis are, so far as is known, without parallel.

(10) APPENDICES

(10.1) Impulse Response of Ladder Networks

Although the maximally-flat and equal-ripple ladder networks are well known, explicit expressions for impulse response as required in the calculation of modem performance are not readily available. The expressions given below were derived for the computation of Tables 2 and 3.

The impulse response for the maximally-flat network of n elements is

$$\left.\begin{aligned} A_n(t) &= \frac{2R_1\omega_c}{n}\sum_{r=0}^{(n/2)-1} F_r(t) \qquad n \text{ even} \\ &= \frac{2R_1\omega_c}{n}\left[\frac{\varepsilon^{-\omega_c t}}{2} + \sum_{r=0}^{(n/2)-(3/2)} F_r(t)\right] n \text{ odd} \end{aligned}\right\} \quad (66)$$

where $F_r(t) = \varepsilon^{-\omega_c t \sin\theta_r} \sin(\omega_c t\cos\theta_r + \theta_r)$

and, for the equal-ripple network,

$$A_n(t) = \frac{2R_1\omega_c}{n}\sqrt{(1+k^2)}\sum_{r=0}^{(n/2)-1} F_r(t) \qquad n \text{ even}$$

$$= \frac{2R_1\omega_c}{n}\frac{1}{\sqrt{(1+k^2)}}\left[\frac{\varepsilon^{-\omega_c t\tanh\beta}}{2} + \sum_{r=0}^{(n/2)-(3/2)} F_r(t)\right] n \text{ odd}$$

where

$$F_r(t) = \varepsilon^{-\omega_c t\sin\theta_r \tanh\beta}$$

$$\left[\frac{\sin(\omega_c t\cos\theta_r + \theta_r)}{1+\varepsilon^{-2\beta}} - \frac{\sin(\omega_c t\cos\theta_r - \theta_r)}{1+\varepsilon^{2\beta}}\right] \quad (67)$$

The potentials at sampling times mt_1, as tabulated in Section 6.3, are calculated by inserting in these expressions $\omega_c t = m\pi$.

(10.2) Filters with Skew-Symmetric Cut-Off

It may be shown that any low-pass filter whose real component of input impedance (equal to the transfer modulus squared if the filter is dissipationless) has skew symmetry about a cut-off frequency ω_c exhibits an oscillatory impulse response which passes through zero at $t = m\pi/\omega$. The derivation was suggested by a parallel procedure due to Sunde.[12]

The function $R(\omega)$ can be split into two parts, respectively rectangular and skew-symmetric, as shown in Fig. 25. The latter part is conveniently written as a function $R_2(u)$ of a variable $u = \omega - \omega_c$ and is an odd function. By use of the Fourier integral (48) the impulse response may be obtained in the form

$$A(t) = \frac{2}{\pi}\sin\omega_c t\left[\frac{1}{t} - 2\int_0^{\omega_c} R_2(u)\sin utdu\right] \quad . \quad (68)$$

Three properties of the filter can be deduced immediately from this equation:

(*a*) Like the rectangular filter, it gives zero potential at times $t = n\pi/\omega_c$ in response to an impulse, so that its pulse-sequence impedance to a pulse train at frequency $2\omega_c$ is a constant.

(*b*) The integral vanishes for $t = 0$, so that $A(0)$ is independent of $R_2(u)$ and hence the pulse-sequence impedance is the same constant for all such filters. This agrees with the fact that the resistance integral, and hence the shunt capacitance, is the same.

(*c*) The integral is positive for $R_2(u)$ of the general form shown in Fig. 23: consequently the oscillatory impulse response is smaller in magnitude for a filter with gradual cut-off than with infinitely sharp cut-off, which is in agreement with common experience. Small variations in alignment of filters or timing of samples should cause less variation in pulse-sequence impedance than with a sharp cut-off.

A filter of the type described has lost one unrealizable feature of the rectangular filter, namely the infinitely sharp cut-off: but it it retains another, namely infinite attenuation over a band of frequencies. Any practical approximation will have finite attenuation just outside the band and a defined asymptotic behaviour at high frequencies. We assume that a ladder filter of the type shown in Fig. 24 can be made to give an approximately skew-symmetric transition between zero loss at low frequencies and a loss of (20 log h) decibels at $2\omega_c$, followed by a slope of $6n$ decibels per octave. An extension of the preceding argument shows that the region below $2\omega_c$ contributes nothing to the impulse response at $t = m\pi/\omega_c$, so that to find these spot values we need consider only the tail of the curve:

$$A\left(\frac{m\pi}{\omega_c}\right) = \frac{2}{\pi h^2}\int_{2\omega_c}^{\infty}\left(\frac{2\omega_c}{\omega}\right)^{2n}\cos\left(\frac{m\pi\omega}{\omega_c}\right)d\omega \quad . \quad . \quad . \quad . \quad (80)$$

$$\simeq (-1)^n\frac{4\omega_c}{\pi h^2}\left[\frac{2n}{(2m\pi)^2} - \frac{2n(2n+1)(2n+2)}{(2m\pi)^4} + \ldots\right]$$

The asymptotic expression converges only for $m\pi > n$, but in all cases the first term is an upper bound.

For likely values of h and n this is very small for all m, except perhaps 1. As an example, take a 4-element filter whose attenuation at $2\omega_c$ is the same as that of a maximally-flat filter ($n = 4$, $h^2 = 257$). In terms of the initial response $A(0)$, the response at the next sampling instant ($m = 1$) is about 0·001 5, at the second ($m = 2$) about 0·000 3, and substantially smaller for $m \geqslant 3$. These values are all smaller (and except for the first they are some orders of magnitude smaller) than those obtained with the known filters for which Tables 2 and 3 were computed. We conclude that a realizable approximation could exhibit substantially the desired property, that the overall transmission of a pulse system should equal the transfer function of the filters alone.

[The discussion on the above paper will be found on page 479.]

621.376.53 : 621.395.4

The Institution of Electrical Engineers
Paper No. 2475 R
Dec. 1957

EFFICIENCY AND RECIPROCITY IN PULSE-AMPLITUDE MODULATION: PART 2—TESTING AND APPLICATIONS

By J. C. PRICE, B.Sc.

(*The paper was first received* 2*nd May, and in revised form* 21*st October*, 1957. *It was published in December*, 1957, *and was read before the* RADIO AND TELECOMMUNICATION SECTION 19*th March*, 1958.)

SUMMARY

The paper is the second of two papers on efficiency and reciprocity in pulse-amplitude modulation, and experimental and practical aspects are discussed. Transmission with an overall loss of about 2 dB is achievable in practice and should be capable of improvement. Response is level up to about 4 kc/s using 10 kc/s sampling, depending on the choice of filter. Subjective speech tests show that satisfactory transmission can be obtained with an economical filter network. Advantage is taken of the 'transparency' of the system to provide voice-frequency signalling. The potential applications for the technique of efficient and reciprocal pulse-amplitude modulation are reviewed briefly.

(1) INTRODUCTION

It has been found possible to achieve an efficient and reciprocal method of pulse-amplitude modulation for multiplex applications by use of storage reactances. A comprehensive mathematical exposition is set out in the companion paper by K. W. Cattermole.[1] In the present paper the experimental evidence supporting the ideas is dealt with, and an account is given of the way in which some of the practical problems have been met.

The method consists essentially of the use of a storage reactance at each line termination. The store at the transmitting end is charged slowly and continuously from the audio signal. It is then discharged in a pulse to a common multiplex busbar. The receiving store charges up quickly during the pulse and discharges slowly and continuously into the audio load. Objective tests on several types of storage network are discussed in the following Section. The subjective tests have been grouped together in Section 3.

(2) OBJECTIVE MEASUREMENTS

(2.1) Purpose

A variety of storage networks were constructed, so as to obtain experimental confirmation of the theory, to observe departures in behaviour from that of the ideal circuit due to practical imperfections such as switch loss, and to assist the choice of a suitable network for an experimental electronic switching model.

Theoretical considerations suggest that a transmission response approximately level to 4 kc/s (with 10 kc/s sampling) may be achieved with a suitable circuit. The results of this work support the theory and show that it is possible to devise a relatively simple network which gives adequate transmission level and frequency response for likely applications.

(2.2) Choice and Construction of Filters

It has been stated[1] that the ideal filter network for storage has an impulse response of form $\dfrac{\sin(t/t_1)}{t/t_1}$, where t_1 is the repetition period of pulses. The corresponding transmission should be approximately level up to frequencies near 5 kc/s and then fall off as rapidly as possible. An ideal filter frequency response is shown in Fig. 1(*d*), where f_c is the cut-off frequency, together with progressive approximations in Fig. 1(*a*)–(*c*).

The filters were designed as maximally-flat ladder filters for use with one end open-circuited.[2] The terminating capacitance

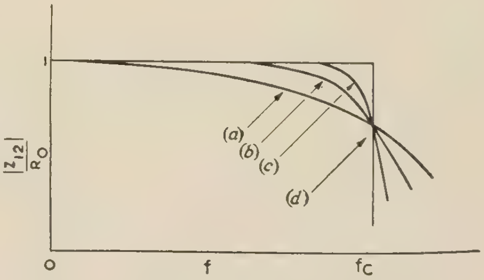


Fig. 1.—Filter frequency-response curves.

(*a*), (*b*) and (*c*) are successive approximations to the ideal, (*d*).
f_c = Cut-off frequency.

consisted of that of the delay line or tuned circuit, in these experiments fixed at 2 000 pF. For the tests a pair of 12-element filters were constructed to approximate closely to optimum conditions. A pair of 4-element filters were constructed as approximating to what was anticipated to be a practical compromise between performance and cost of components. Other more simple filters were also tested. The capacitors were close-tolerance silvered-mica types. The inductors were constructed using ferrite cores with a mica-spaced gap. Typical Q-factors were of the order of 50 or 60.

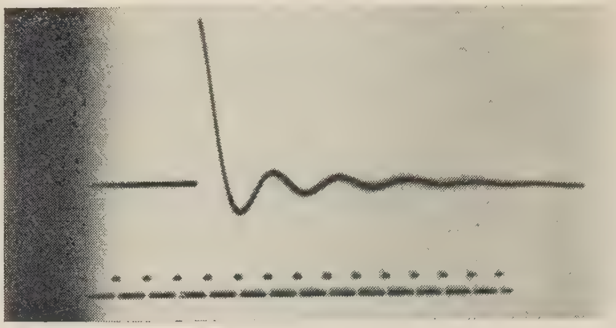


Fig. 2.—Impulse response for 12-element filter.

A photographed impulse response is shown in Fig. 2 for a 12-element filter. The calibration curve shows 25 microsec pulses at a recurrence frequency of 10 kc/s. Frequency-response characteristics for the 4-element and 12-element filters are plotted in Fig. 3. These show some departure from theoretical curves for lossless components in having a rather lower cut-off frequency and some ripple in the case of the 12-element filter. However,

Mr. Price is with Standard Telecommunication Laboratories Ltd.

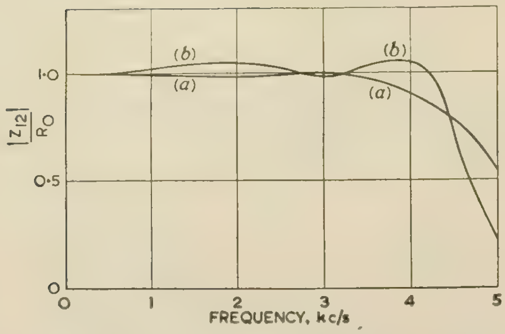


Fig. 3.—Measured frequency response of filters.

(a) Four-element filter.
(b) Twelve-element filter.

it is found that small variations in filter characteristics do not critically affect pulse transmissions.

In later experiments, therefore, equal-ripple (Chebŷshev) filters of small deviation have been used in preference to maximally-flat filters. For a given bandwidth and number of elements these have a sharper cut-off rate, which is useful in eliminating the unwanted components between 6 and 10 kc/s. This feature is especially valuable in the voice-frequency dialling arrangement mentioned in Section 4. The frequency responses of two such filters are shown in Fig. 4. These are, respectively, a 4-element filter with 2% ripple and a 5-element filter with 10% ripple.

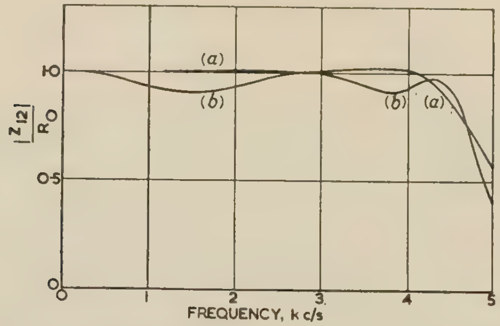


Fig. 4.—Measured frequency response of Chebŷshev filters.

(a) Four-element filter (2% ripple).
(b) Five-element filter (10% ripple).

(2.3) Transmission Measurements

A basic circuit for transmission measurements is shown in Fig. 5. Switches S_1 and S_2 are ganged so that in the upper

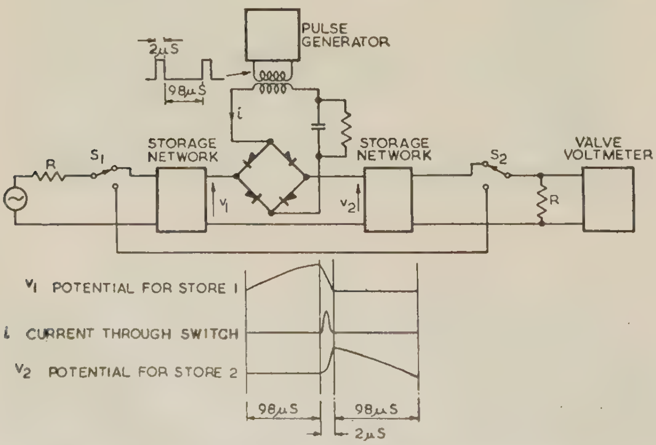


Fig. 5.—Basic circuit for transmission measurements.

position transmission is through the modem system and, in the lower position provided for comparison, transmission is by direct connection.

The speed of operation was chosen to be in accordance with practical requirements for good telephone quality with perhaps 25 channels. Thus the diode bridge gate between modems was operated at a repetition rate of 10 kc/s by pulses of 2 microsec duration.

(2.3.1) Transmission without Filters.

The first transmission measurements were made between impedances chosen to give maximum transmission at zero frequency without filters. The condition for this was shown in Reference 1 to be $\alpha = 1 \cdot 25$, where $\alpha = t_1/(CR)$, t_1 is the repetition time, C the storage capacitance, and R the resistance of source and load.

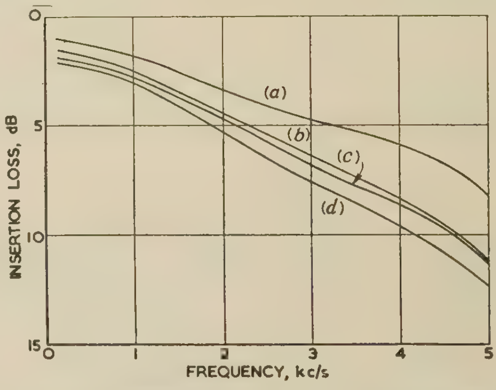


Fig. 6.—Frequency response of system (no filters, minimum loss at direct current).

(a) Theoretical curve for lossless store and switches.
(b) Measured curve for tuned-circuit storage.
(c) Measured curve for delay-line storage.
(d) As (c) but with additional switch resistance (single diodes in bridge).

In Fig. 6 curve (a) shows the theoretical curve for a lossless store and switches of zero forward impedance. Curve (b) shows the measured insertion loss for tuned-circuit stores, these being series circuits tuned to 250 kc/s. Curve (c) shows the measured insertion loss for 1 microsec delay lines. In both these tests the gate impedance was made as low as possible by the use of 3 point-contact diodes (type 2X/105G) in parallel in each arm. The storage time and capacitance of available junction diodes prohibited their use. Curve (d) is a repetition of curve (c), with single diodes in the arms of the bridge. The 3-diode arrangement would not be favoured in practice, on the grounds of the additional cost and increased crosstalk due to reverse leakage.

It will be seen that the experimental curves are similar in shape to the lossless theoretical one but exhibit greater attenuation increasing with frequency. An analysis of the use of a delay line with a switch of appreciable resistance shows that an increase of attenuation with frequency is to be expected, although of rather less extent than that observed. It is possible a more detailed consideration of types of losses involved would lead to a closer agreement with measured values. Fortunately, however, this discrepancy in frequency response is not observed in the measurements with filters, which are the cases of practical interest.

(2.3.2) Transmission with Filters but without Transformers.

The next tests were made with network elements chosen to give a maximally flat response. Measurements were carried out using a diode bridge switch with three diodes in each of the arms. Curve (a) in Fig. 7 shows the case of maximally flat transmission with the capacitance of the storage device as the only element. Curves (b)–(e) show the effect of increasing the

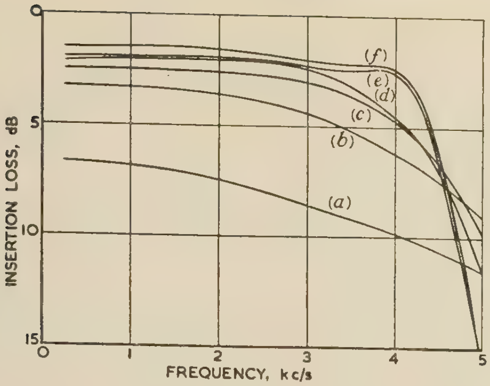


Fig. 7.—Frequency response of system (high-impedance terminations).

(a) One-element filter and delay line.
(b) Two-element filter and delay line.
(c) Three-element filter and delay line.
(d) Four-element filter and delay line.
(e) Twelve-element filter and delay line.
(f) Twelve-element filter and tuned circuit.

number of filter elements. These show the predicted approximations to level response improving with more elaborate filters. Curve (f) shows the transmission where the delay line of previous experiments has been replaced by a tuned circuit. The lower loss is thought to be due to lower losses in the inductance of the tuned circuit compared with those in the delay line.

(2.3.3) Transmission using Filters and Transformers: the Effect of Added Capacitance.

In any practical multiplex system the pulse length t_2 will be much smaller than the sampling period t_1. Hence the impedance of the pulse-forming line, R_2, will be much smaller than that of the signal source and load, R_1, by the approximate relation $R_1/R_2 = t_1/t_2$. With $t_2 = 2$ microsec and $t_1 = 100$ microsec there will be an impedance ratio of 50. Choice of line impedance is restricted mainly by the consideration that it must be large compared with the impedance of the diode switch closed but small compared with it open. With existing diodes 500 ohms was chosen for line impedance. Therefore a source and load impedance of about 25 kilohms is required, and to provide this with 600-ohm audio terminations necessitates the use of transformers.

The transformers used in the tests were made with a turns ratio of 1/6·4, and the last element of the filter was adjusted to compensate for the transformer leakage inductance. Transmission obtained with the 12-element filters is shown in Fig. 8. Curves (a) and (b), respectively, show the response for a tuned circuit and a delay line with a single gate. Curve (c) shows the case where transmission is via a common rail isolated from two stores and earthed between pulses. For curve (d) 1 330 pF is connected on the common line. It is shown in Reference 1 that this corresponds to a condition of theoretically zero loss and may be useful in practice where many channels on the common rail contribute to a high capacitance. As was predicted, the variation in level due to small changes in timing is reduced by this arrangement.

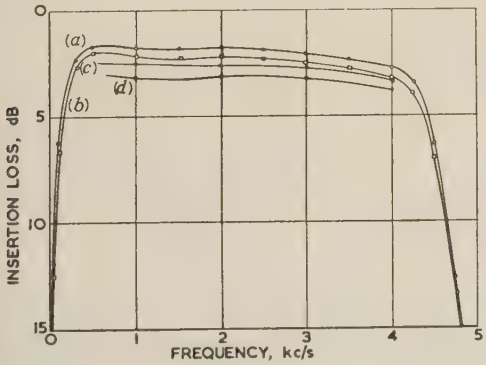


Fig. 8.—Frequency response of system (600-ohm terminations using transformers).

See Fig. 12 for nomenclature of gating arrangements.

(a) Twelve-element filter: single gate (α), tuned circuit.
(b) Twelve-element filter: single gate (α), delay line.
(c) Twelve-element filter: triple gate (α, β, γ), tuned circuit.
(d) Twelve-element filter: triple gate (α, β, γ), tuned circuit. 1 330 pF on common line.

Curves for the 4- and 5-element equal-ripple filters are given in Fig. 9. It will be seen that the 4-element filter with 2% deviation gave a level response to nearly 4 kc/s, but rejection outside the pass band was not much better than for a maximally flat filter, and probably not good enough for the voice-frequency dialling arrangement; the 5-element filter with 10% deviation obtained adequate rejection at the price of irregularity in the pass band. Something intermediate with five elements and about 5% ripple would probably be satisfactory for all purposes.

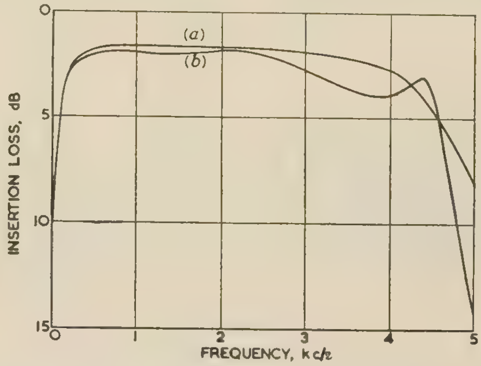


Fig. 9.—Frequency response of system (600-ohm terminations using transformers).

(a) Four-element Chebyshev: triple gate, tuned circuit.
(b) Five-element Chebyshev: single gate, tuned circuit.

(2.4) Waveforms

A photographic record was taken of several waveforms commonly encountered in high-efficiency pulse modulation and demodulation using delay lines. Fig. 10 shows a typical waveform on the multiplex output rail of a 24-channel transmission terminal using this technique: (a) is a 6 microsec 10 kc/s synchronizing pulse; (b) is a 2 microsec 125 kc/s fine synchronizing pulse; (c) is a modulated audio pulse. Both modulated and unmodulated audio channels are shown.

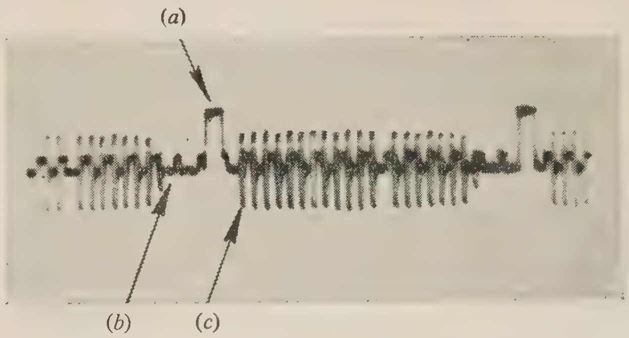


Fig. 10.—Time-division-multiplex output waveform.

(a) 6 microsec 10 kc/s synchronizing pulse.
(b) 2 microsec 125 kc/s fine synchronizing pulse.
(c) Modulated 2 microsec audio pulse.

Fig. 11(*a*) shows the voltage-modulation envelope on a transmitting modem delay-line store in conjunction with a 4-element filter. The repetition period in this exposure was 114 microsec. The gradual charging up and characteristic delay-line discharge will be noted. This is followed by a few fluctuations due to imperfection of switching and delay-line waveforms.

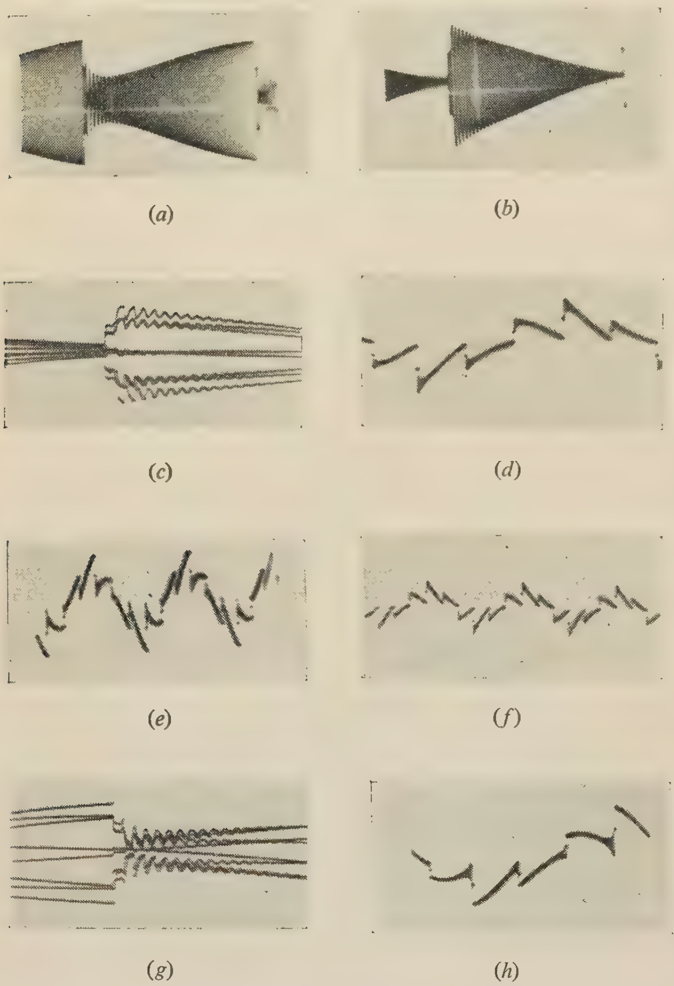


Fig. 11.—Voltage-modulation envelope on transmitting modem.

(*a*) Voltage-modulation envelope on transmitting modem.
(*b*) Voltage-modulation envelope on receiving modem.
(*c*) Voltage waveform on transmitting modem.
(*d*) Voltage waveform on receiving modem.
(*e*) Voltage waveform on transmitting modem over several audio cycles.
(*f*) Voltage waveform on receiving modem over several audio cycles.
(*g*) Discharge of transmitting modem into line.
(*h*) Charge of receiving modem from line.

Fig. 11(*b*) shows the corresponding receiver modem modulation envelope. The delay line is charged up quickly by the pulse and discharged slowly into the audio source. Similar transmitter and receiver waveforms displayed over a longer period are shown in Figs. 11(*c*) and (*d*). A 12-element filter was used here and audio synchronized with pulse repetition frequency.

Figs. 11(*e*) and (*f*) show transmitter and receiver waveforms over several audio cycles. It should be noted that these are the waveforms at the pulse end of the filter, the pulse-frequency ripple in the final output being almost imperceptible.

Fig. 11(*g*) shows in detail the discharge of the transmitter into the line, a 12-element filter being used with 1·25 kc/s audio and 10 kc/s recurrence frequency. Fig. 11(*h*) shows the corresponding conditions at the receiver.

(3) SUBJECTIVE TESTS

(3.1) Purpose

The imperfections of this or any other sampling system include band limitation, leakage of steady tone, and generation of spurious frequencies in the presence of a signal. Their combined effect in degrading transmission is not easily assessed by objective means. While it is possible by choice of a sufficiently elaborate filter or a sufficiently high sampling rate to reduce all these imperfections to a point where they are clearly negligible, it is economically necessary to use large numbers of channels despite the speed limitations of electronic switches, and to keep the apparatus per subscriber to a minimum. To decide how much degradation was tolerable in conjunction with normal telephone apparatus, and thereby to settle on a simple but adequate modem network, subjective speech tests were carried out.

The value of the results is limited by the artificial provisions and small scale of the tests. Nevertheless, an order of preference among various arrangements was registered which accords with objective properties. It was established that the type of imperfection occurring in this system is not unduly offensive even in a rudimentary embodiment: and a reasonable standard was determined for an adequate but economical network.

(3.2) Circuit Conditions

Two telephone subscriber's sets of standard type (No. 332) were connected by alternative paths providing seven types of transmission:

(*a*) Direct connection.
(*b*) Delay-line storage multiplex with no filter.
(*c*) Delay-line storage multiplex with 4-element filter.
(*d*) Delay-line storage multiplex with 12-element filter.
(*e*) Tuned-circuit storage multiplex with no filter.
(*f*) Tuned-circuit storage multiplex with 4-element filter.
(*g*) Tuned-circuit storage multiplex with 12-element filter.

In the course of some tests the effects of changing subscriber's sets and the introduction of capacitance on the line were also tried.

A block diagram of the circuit used for listening tests is shown in Fig. 12. Subscriber's sets are energized by a d.c. supply not shown and fed through an audio transformer to switch S1a. In position 1 of this switch the multiplex system is by-passed via a 2 dB pad, which is included to make the speech level approximately the same as for the multiplex system. On position 2, connection is made to the simplest multiplex system using a source of optimized impedance but no filter. A pulse generator giving a 2 microsec pulse recurring at 10 kc/s is used to operate the diode bridge gates shown. When required, line capacitance can be added at point A to simulate the total capacitance of other channels. Switch S2 makes available either delay-line storage or tuned-circuit storage. Position 3 of switch S1 makes available a 4-element maximally-flat low-pass filter with cut-off approximately half the repetition rate. Position 4 makes available a 12-element filter which has a more even response in the pass band and a sharper cut-off than the 4-element filter.

(3.3) Survey Conditions

Thirty persons, chosen to include both sexes and as much variety of age and occupation as was available, were asked in turn to converse with an interviewer over the telephone link described. The seven conditions were selected at random by the interviewer, the other party having no clue to the type of transmission being used and, in most cases, no preconceived idea of the type of defects likely to occur. After listening and speaking to get the feel of each set of conditions, subjects were asked to award marks according to their judgment of the system as

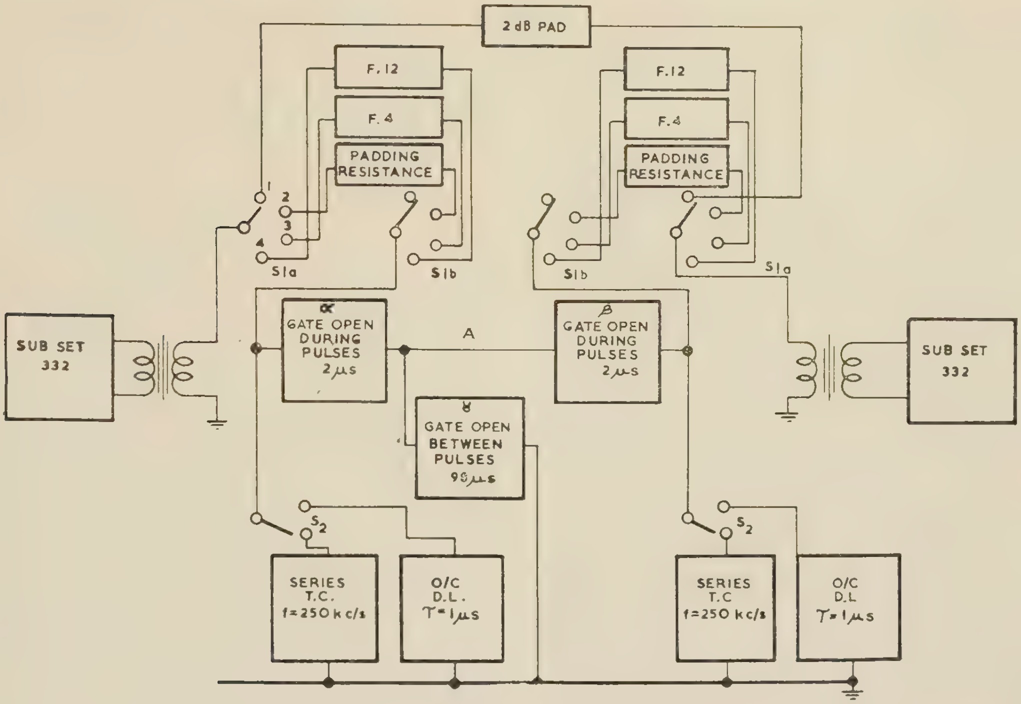


Fig. 12.—Circuit used for listening tests.

compared with normal public telephone intelligibility, according to the following rating:

Opinion of intelligibility	Mark
Considerably above average	5
Above average	4
Average	3
Below average	2
Considerably below average	1

Intermediate marks were permitted if requested by people taking part, and a note was taken of any comments that were volunteered regarding the service. Test conditions were taken in random order.

(3.4) Results

(3.4.1) Average Estimate.

The average estimate of each type was as follows in order of merit:

Direct connection	4·18
12-element filter and delay line ..	4·15
12-element filter and tuned circuit ..	4·03
4-element filter and tuned circuit ..	3·92
4-element filter and delay line	3·48
No filter with tuned circuit	3·02
No filter with delay line	2·80

(3.4.2) Impressions.

The direct connection and the 12-element filter modems were most generally praised, as expected. Each of these was described as the 'best' by one or more subjects. However, only the direct connection educed the most favourable comments on fidelity: the modems met with several mild criticisms, mostly using onomatopoeic terms such as 'buzz', 'burr' or 'mush' to describe a slight crackling or rasping noise most noticeable on explosive consonants. This effect is undoubtedly present; it appears to derive in part from the generation of difference frequencies in the sampling process, but was accentuated by overloading on peaks and could probably have been reduced by limiting. With the 12-element filters it is not very pronounced, and passed unnoticed by some subjects even with loud talking.

The 4-element filter modems exhibit the same effect to a rather greater degree, and similar comments were made. This condition also provoked the interesting observation that the buzzing is more apparent on the sidetone than on transmitted speech. The rudimentary modems without filters, although invariably passing perfectly intelligible speech, were criticized not only for the effect mentioned above but also for whistles and for obvious treble attenuation.

Capacitances of 1 470 pF and 1 330 pF introduced on the common line did not conspicuously alter the general pattern of results except in the case of no filter with the delay line, where slight deterioration was noticed.

It would appear that a discerning ear clearly detects the improvement in reproduction when the response is not restricted to 5 kc/s. However, multiplex systems using 4-element and 12-element filters are quite acceptable as giving good intelligibility as distinct from fidelity. A full-time telephone operator said that all transmissions sounded very good compared with her switchboard. People who habitually used the telephone rated the multiplex systems with filters nearly as highly as the direct connection. People less used to the telephone tended to be more distracted by the imperfections of the less good systems and accordingly marked them down. Some people found it difficult to distinguish much difference between systems. The difference also seemed less apparent when listening to women speaking than to men.

(3.5) Conclusions on Listening Tests

A 4-element filter seems a good provisional choice, judging by speech transmission alone. There is a case for a slightly better filter which would be appreciated by some people. A filter of considerably more elements is unnecessarily extravagant with

components. A slight adjustment of choice of filter may occur in practice, where lines may be subject to distortion, interference and attenuation due to other causes; and the voice-frequency dialling scheme requires a somewhat sharper cut-off than was used in these tests.

It should not be thought that a multiplex system implies a lower quality of transmission. For example, the net band over a universal telephone system used at present may be 400 c/s–2 kc/s or less,[3] with which the bandwidth of the pulse modem used in these tests compares most favourably.

(4) VOICE-FREQUENCY SIGNALLING

The theory of 2-way pulse modems developed in Reference 1 deals specifically with the problem of obtaining maximum power transfer, which necessarily leads to circuits which match their terminations fairly well. The fact that the system is both low-loss and reciprocal implies, however, that it is 'transparent' and that impedance changes can be 'seen' through it. In less visual language, the current incident on a mismatch can be considered, in the usual way, as dividing into a transmitted and reflected component. If the low-frequency termination of a pulse modem differs appreciably from the matching impedance, then some of the signal current is reflected back into the modem: because of the reciprocity, it is translated into pulses and conveyed through the normal route back to the signal source, and the input impedance of the remote modem, as faced by the signal source, is modified as a result of the mismatch on the output. This viewpoint is familiar in transmission-line theory, and is readily developed in detail for the pulse system.

The application of reflections to voice-frequency signalling is illustrated by the experimental circuit shown in Fig. 13. Two

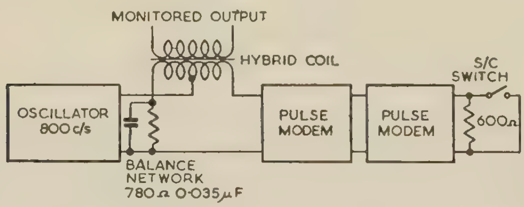


Fig. 13.—Basic circuit for voice-frequency signalling with two synchronous pulse modems.

synchronous pulse modems were connected together, one having a matched load, with provision for short-circuiting, the other being fed with a v.f. tone through a hybrid transformer, with a balancing network on the conjugate arm chosen so as to annul the output from the balancing winding. When the short-circuit was applied to the remote modem, a substantial output was obtained from the balanced winding. Since the input impedance of the modem is not a constant quantity, but varies periodically, it could not be exactly balanced by a passive network: the v.f. component of the output could be annulled, but some higher-frequency components remained. Despite this, the ratio of the peak amplitudes in the two conditions (short-circuit and terminated) exceeded 20 with practical modems, a factor which is readily detected with a fair margin of safety. A fairly sharp cut-off is required in the low-pass filter to obtain this figure: the 12-element maximally-flat and the 5-element 10% ripple filters described in Section 2 were adequate, but the simpler filters were not.

The potential on the receiving store is normally of the form shown in Fig. 14(*a*). When the short-circuit is applied the waveform changes to that shown in Fig. 14(*b*), with approximately equal weight on either side of zero. It is clear that the waveform (*a*) contains a substantial low-frequency component, whereas (*b*)

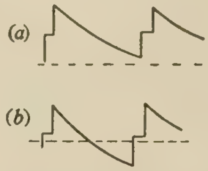


Fig. 14.—Voltage on receiving modem store.
(*a*) Output correctly terminated.
(*b*) Output short-circuited.

does not: this would be expected, since little or no power is delivered into the short-circuited modem.

In telephone-switching systems, pulse signals produced by a dial are used either to actuate a selecting mechanism directly, or to set up selecting information in a register. They must be conveyed from any calling subscriber to the unit which utilizes them, and to avoid the provision of further switched paths it is useful if they can be sent through the speech path. The 2-way pulse modem, like certain other electronic switching schemes, is normally used with transformers to obtain reasonable impedance levels at all points: it cannot pass direct current or sub-audio frequencies efficiently. Consequently, dialling signals can be conveyed through it only after translation to voice frequency. Clearly, it is uneconomical to provide a translator or generator for every subscriber's line. Using the effect described, a central tone generator and detector, associated with a register, can be connected through the speech path to any calling subscriber and tone from the generator reflected to the detector by applying a short-circuit, or other substantial mismatch, on the subscriber's line circuit.

The reflection could occur at the subscriber's set, but to avoid the effect of varying line lengths and attenuation it is better to use d.c. signalling on the subscriber's loop and to cause a v.f. mismatch at the exchange. The mismatch must be applied, by economical means, when the subscriber's line current is either interrupted (as in the conventional system) or increased (as is more convenient in an electronic system under development). Quite a good short-circuit can be obtained using two semiconductor diodes (preferably low-impedance junction diodes) which are normally biased off but arranged (in opposite senses) to form a conducting path when required. A single diode shunted across the line has been found fairly effective and is easily arranged to conduct or not according to the magnitude of the line current.

A preferred arrangement is to use a saturable transformer in the audio path. This produces almost as good a reflection as the short-circuit, and a better one than with a single diode: the fact that its phase shift is different is not significant in this application. Since the transformer must not approach saturation on normal line current, or distort speech signals, it is best to make the line current drawn during a dial pulse substantially larger. A refinement which assists in obtaining a clear-cut distinction between the two conditions is to balance the normal line current by a direct current in the other winding, as shown in Fig. 15. In a circuit designed for use with a low-consumption subscriber's set developed in this laboratory, the current magni-

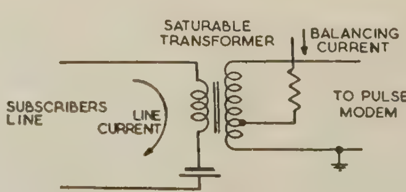


Fig. 15.—Short-circuiting by means of a saturable transformer.

tudes are: normal line current 4 mA (± 1 mA); balancing current 4 mA; dialling current 12 mA (± 2 mA). No difficulty was found in making a suitable transformer.

Further tests were made using a standard telephone dial, completing a resistive circuit, to drive pulses of current through a saturable transformer as above, with an amplifying detector using

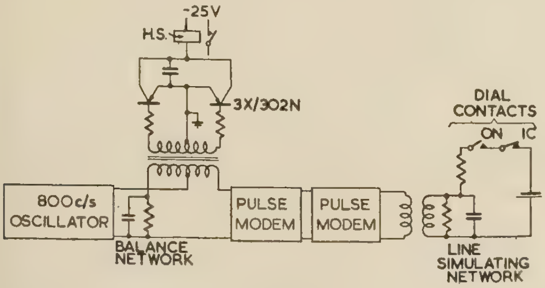


Fig. 16.—Circuit for dialling over a modem system.

two junction transistors (Fig. 16) to operate a relay. With precise balancing, the arrangement functioned as expected. In a practical telephone network, the impedances of the subscribers' lines vary considerably and exact balance will not be possible. To see how much trouble this would introduce, the line-simulating network (Fig. 16) was set at 1 100 ohms and $0\cdot2\,\mu$F (taken rather arbitrarily as a typical value) and the balancing network adjusted. It was found possible to set the detector sensitivity so that correct operation occurred over a wide range of values of the line-simulating network, namely 400–1 400 ohms and $0\cdot1$–$0\cdot3\,\mu$F; although the margin of safety was, of course, small near the extreme values. It is not possible to say whether the range of impedances encountered in normal telephone practices could be covered, since this range is not known. However, it seems quite safe to say that with short lines, such as are connected to a p.b.x., the method should be applicable.

(5) ELECTRONIC SWITCHES

(5.1) Switching Devices

It is not proposed to describe in detail the transistor and diode circuits in conjunction with which the pulse modem was developed.[4] Although the principles can in fact be used with mechanical, thermionic, semi-conductor or magnetic switching devices, the semi-conductors appear most suitable, combining moderately high speed, very low forward potential difference when conducting, and reverse leakage when non-conducting which is sufficiently low for most purposes.

The major part of the work was done with balanced switches of the form shown in Fig. 17 using four point-contact diodes. For switches in continual periodic operation, dynamic bias developed across a condenser and resistance suffices: otherwise

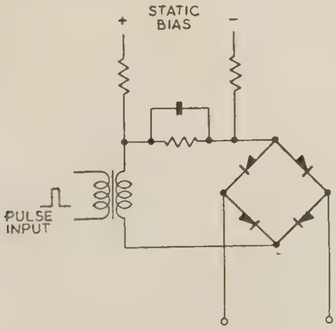


Fig. 17.—Balanced switch using point-contact diodes.

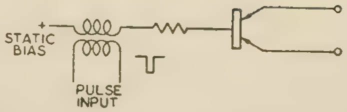


Fig. 18.—Symmetrical junction switch.

some static bias is also provided. More recently, symmetrical junction transistors have been available (Fig. 18); these consume less switching power and have lower forward drop than diodes, and will be favoured in future work. Switching pulse trains (usually 2 microsec pulses at 10 kc/s) have been derived from blocking-oscillator circuits using point-contact transistors.

(5.2) Switching Power

Electronic switching devices also impose voltage and current limits. With balanced diode switches coupled by transformers to transistor pulse generators, there are two distinct types of limit—on the voltage and on the current separately, due to the diodes, and on their product, due to the transistor. The current rating required is set by the peak current during the pulse; the voltage rating by the peak p.d. to be isolated in a multiplex system, i.e. the sum of the peak voltage across the store between pulses and that on the common line during a pulse. These quantities below are tabulated for unit p.d. transferred from one store to another for each type of store mentioned:

Store	*Current**	*Voltage*	*Product**
Delay line	1	$1\cdot5$	$1\cdot5$
LC without line capacitance	$1\cdot57$	$1\cdot5$	$2\cdot36$
LC with line capacitance .. ($C' = 2C/3$)	$2\cdot04$	$1\cdot75$	$3\cdot57$

* In units of C/T, where C is storage capacitance and T is pulse duration.

At present, the pulse power available from a transistor is only just enough to drive a diode switch in a speech modem, so that the extra power required by the LC circuit is a temporary hindrance: however, the symmetrical junction transistor is easily applicable in any of the circuit arrangements.

(5.3) Losses and Crosstalk

The minimum loss so far achieved is a little under 2 dB, of which rather more than half is due to dissipation and other departures from the ideal in the storage networks, and the remainder due to losses in the switches. Since the bulk of the work was done, diodes have been received in sample quantities which allow some reduction in loss, and the junction transistors are also somewhat better: with foreseeable development in semi-conductor devices most of the present switch loss should be eliminated.

Storage network imperfections may be classified as:

(a) Loss in pulse-forming network.
(b) Loss in filter coils.
(c) Departure of filter characteristic from the ideal.

The first two might be reduced slightly by component developments: in the low-frequency filter coils there will be an economic as well as a technical limit. With a single LC circuit, (a) is normally a small fraction only of the total loss. The effect of the third factor can be reduced by use of more elaborate filters, but, apart from obvious drawbacks, the addition of further coils increases the filter dissipation: it is doubtful if anything is gained by going beyond six to eight elements. All things considered, the minimum attainable loss is probably about 1 dB or a little over.

Crosstalk in a multiplex system depends on the reverse leakage and capacitance of the switching device. It will be noticed that only one series switch has been used, whereas it is common

practice in gating circuits to use several series and shunt diodes in cascade. Although this would be feasible in pulse modems, it is not necessary because any switching device provides better crosstalk margins when used in a modem than when used as a simple sampling gate. There are two chief sources of crosstalk, each of which is attenuated by use of storage:

(*a*) Leakage of unwanted audio signals to line at the transmitter during pulse period of wanted channel. Because the impedance ratio R_1/R_2 is quite large ($\simeq t_1/t_2$) the impedance R_2 is much lower [by a factor of the order of $(t_2/t_1)^{\frac{1}{2}}$] than a normal gate load: this impedance acts as the lower part of a potential divider whose upper part is the reverse impedance of the switch, and so increases the voltage attenuation by the factor quoted.

(*b*) Leakage of unwanted pulse to the channel circuit at the receiver. The audio component of the unwanted pulse is extracted, whereas the wanted channel pulse is completely utilized by the process of efficient reception.

From each cause separately, and hence from their sum, the efficient modem is better than a simple gate by a factor of about t_1/t_2 in power: thus, with the same potentials to be isolated, the crosstalk margin would be better by about $10 \log t_1/t_2$ decibels. However, the potential on the storage capacitance changes continually over the sampling cycle, and is a little over double the pulse amplitude just before the sampling pulse. Consequently, crosstalk from a given channel into the preceding channel is about 4 dB more than appears from the above reasoning, and crosstalk into the following channel very much less. With 2 microsec pulses at 10 kc/s (25-channel rate) measured values of crosstalk margin were in the range 46–66 dB with commercial point-contact diodes, and 55 to over 70 dB with junction transistors or improved laboratory diodes.

(6) POTENTIAL APPLICATIONS

There would appear to be three main fields of application for the technique which has been evolved for obtaining efficient and reciprocal pulse-amplitude modulation (p.a.m.). These are:

(*a*) Direct use of p.a.m. in transmission systems.
(*b*) Use as an intermediate stage towards other pulse methods.
(*c*) Use in switching systems.

(6.1) Transmission

It is generally accepted that pulse methods are not suited to long-distance transmission on cable, since they require unusually good phase equalization and consume a grossly uneconomical bandwidth. Pulse methods are more readily used over radio links. Pulse-amplitude modulation is not commonly used because it is supposed to be more vulnerable to noise than are other modulation methods. In fact, much depends on the method of carrier modulation. Landon, in a comprehensive study of multiplexing,[5] concludes that with a given bandwidth, p.a.m. on frequency-modulation is somewhat superior to p.t.m. on amplitude-modulation in rejection of both thermal and impulsive noise, and causes less interference with adjacent channels, provided that both positive and negative pulses are used. The radio link is not reciprocal, so the terminal apparatus must be worked on a 4-wire basis. The power efficiency of the channelling apparatus used keeps the size, cost and power consumption down, and enables low-power devices such as point-contact transistors to be used.

Other pulse methods commonly require more complex modulators and demodulators; the use of p.a.m. for channelling, together with a single modulation convertor, may often be economical, and also allows the insertion of a common compandor. In this case also the present technique makes for economy.

(6.2.) Switching

In a switching centre, where the multiplex signal is not degraded by a long, noisy and band-limited transmission path, noise and bandwidth are relatively unimportant: economy and reliability, combined with reasonable crosstalk margins, are paramount. For this reason, all the known working models of, or serious proposals for, multiplex switching schemes utilize p.a.m. The economy and simplicity inherent in p.a.m., as compared with other modulation methods, are both enhanced by the use of efficiency and reciprocity. The possibility of passing a signal through several multiplex selectors in cascade, with simple apparatus and low loss, and without converting to 4-wire working, taken in conjunction with an economical means of generating the control pulses and storing the connecting information, may well effect a revolution in the field of electronic switching.

(7) ACKNOWLEDGMENTS

The author is indebted to Messrs. H. Grayson, K. W. Cattermole and R. B. Herman for discussion and suggestions in connection with the work, and thanks Standard Telecommunication Laboratories Ltd. for permission to publish the paper.

(8) REFERENCES

(1) CATTERMOLE, K. W.: 'Efficiency and Reciprocity in Pulse-Amplitude Modulation: Part 1—Principles' (see page 449).
(2) ORCHARD, H. J.: 'Formulae for Ladder Filters', *Wireless Engineer*, January, 1953, **30**, p. 3.
(3) HORTON, A. W., JR., and VAUGHAN, H. E.: 'Transmission of Digital Information over Telephone Circuits', *Bell System Technical Journal*, 1955, **34**, No. 3, p. 511.
(4) CATTERMOLE, K. W.: 'Transistor Pulse Generators for Time-Division Multiplex' (see next page).
(5) LANDON, VERNON D.: 'Theoretical Analysis of Various Systems of Multiplex Transmission', *RCA Review*, 1948, **9**, pp. 287–251 and 433–483.

[The discussion on the above paper will be found on page 479.]

621.373.43 : 621.395.43

The Institution of Electrical Engineers
Paper No. 2577 R
Mar. 1958

TRANSISTOR PULSE GENERATORS FOR TIME-DIVISION MULTIPLEX

By K. W. CATTERMOLE, B.Sc.

(*The paper was first received* 2nd *May, and in revised form* 7th *December*, 1957. *It was published in March*, 1958, *and was read before the* RADIO AND TELECOMMUNICATION SECTION 19*th March*, 1958.)

SUMMARY

Point-contact transistor circuits to generate pulses in the microsecond range are described, together with means of frequency-dividing and interlacing pulse trains and their application to time-division operation of telephone transmission and switching systems.

(1) INTRODUCTION

Time-division multiplexing of telephone circuits is well established in radio transmission and is now considered for exchange switching and cable transmission also. Whatever the modulation method, numerous pulse trains are required at reasonably precise relative time positions. The main functions are:

(*a*) Generation of steady timing pulse trains.
(*b*) Frequency division of pulse trains.
(*c*) Interlacing of many trains with permanent time positions.
(*d*) Interlacing of many trains with assignable time positions.

Transistor circuits have been developed for each function. While either junction or point-contact transistors can be used, the point-contact type is more immediately suitable for the class of circuit considered and most of the paper describes its use.

For normal telephony, sampling rates of 8–10 kc/s are used. Transmission groups of 12 or 24 channels are commonly used and the equivalent of at least one extra channel is required for synchronizing. Switching selectors require at least 16–18 positions, preferably more. The initial aim in this work was the establishment of basic techniques for a pulse repetition rate of 250 kc/s, equivalent to 25 channels with 10 kc/s sampling.

(2) TRANSISTOR PULSE GENERATORS

(2.1) A Basic Circuit

The first necessity is a basic pulse generator, capable of producing at sufficiently high speed a train of approximately rectangular pulses whose width shall depend on the external circuit rather than on the device.

A distinctive property of the point-contact transistor is that it has current gain without phase change; this means that any simple back-coupling element, such as an impedance in series with the base or an admittance between emitter and collector, provides regenerative feedback which can be used as the basis of an oscillator or trigger circuit. To produce pulses of defined width the coupling element must define a time, and hence should contain a delay line or a tuned circuit.

A suitable circuit is shown in Fig. 1(*a*). The collector of the transistor is connected, through a load resistance of a few kilohms, to a negative supply (in the range 20–50 volts), and the emitter, through a higher resistance, to a positive supply (capable of delivering a current in the range 1–10 mA). Feedback from collector to emitter is through a delay line used as a 2-terminal admittance. The impedance of the line, in relation to that of the load, is determined by the method used in the Appendix.

Consider the circuit to be quiescent, with the emitter slightly

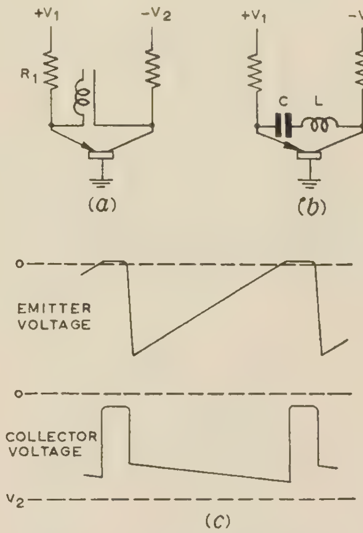


Fig. 1.—Simple pulse generators.

(*a*) With delay-line timing.
(*b*) With *LC* timing.
(*c*) Waveforms.

negative, the collector at a large negative potential, and the line charged to their potential difference. Current from the positive supply flowing through R_1 causes the emitter potential to rise: when it becomes positive, collector current flows, which tends to discharge the line and drive the emitter further positive. The action is regenerative, and continues until the collector potential has risen almost to zero; it produces the leading edge of the output pulse shown in Fig. 1(*c*), and at the same time a step is launched down the line.

An equilibrium condition with steady currents in the emitter and collector, and at the terminals of the line, is maintained for a period equal to twice the line's delay time; then the step of voltage on the line, having been reflected in phase at the open-circuit end, returns to oppose the applied voltage step and greatly reduce current feedback to the emitter. The transistor turns off regeneratively. The collector potential falls rapidly nearly to its initial value; since the reverse resistance of the emitter is high, the line does not discharge appreciably during the trailing edge, and the emitter potential falls to nearly twice the negative potential of the collector.

Current from the positive supply flows into the line, recharging it in the original sense; the emitter potential rises, ideally in a staircase waveform but practically (if, as is common, the recovery time is much greater than the pulse length) in a smooth curve with a few ripples. When the emitter goes positive, the transistor turns on again, and the cycle repeats.

Another suitable circuit is shown in Fig. 1(*b*), in which the delay line has been replaced by a tuned circuit. The operation is generally similar: the transistor turns off regeneratively when current through the tuned circuit falls below that necessary to

Mr. Cattermole is with Standard Telecommunication Laboratories Ltd.

maintain the 'on' condition. The pulse length is about half a period, i.e. $\pi\sqrt{(LC)}$. The rate of rise of the pulse current is limited by the inductance L: an approximate mathematical analysis shows that, if the delay in the transistor were negligible, the rise would be almost exponential, with time-constant $L/(\alpha - 1)R_c$ (where R_c is equivalent to the transistor output resistance in parallel with R_2). In fact, the rate of rise is also limited by the transistor properties; quite commonly these dominate, and little difference is noticed between the waveforms of the two circuits.

These circuits must be distinguished from that with feedback through a condenser only. The latter has no time determination built into the external circuit; it depends much more on transistor properties. For this reason, it has been found useful in transistor testing but not in working circuits.

(2.2) A Triggered Pulse Generator

The free-running pulse generators shown in Fig. 1 are easily converted to triggered operation. Negative bias applied to the emitter through a diode [Fig. 2(*a*)] prevents the transistor from

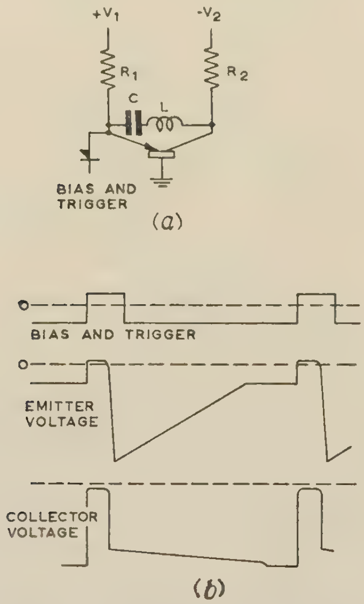


Fig. 2.—Triggered pulse generator.

(*a*) Circuit.
(*b*) Waveforms.

firing until either a positive pulse is applied in series with the emitter bias [Fig. 2(*b*)] or a negative pulse is applied to the base. The recovery time, which depends mainly on CR_1, must be shorter than the interval between trigger pulses. During the recovery period current is flowing into the condenser C through R_1 and R_2: consequently, the emitter and collector potentials vary in opposition, but since R_2 is normally much less than R_1 the excursion at the collector within this period is much smaller.

This circuit gives a positive-going output for a positive-going trigger. By inserting a resistor R_3 in series with the base, a negative-going trigger pulse can be used and a negative-going output pulse obtained. The circuit and waveforms are shown in Fig. 3. In the quiescent state, the base is at negative potential, owing to collector current flowing through the base resistance; the emitter must be biased still further negative to hold the transistor off. If a stable triggering level is required or if only small trigger pulses are available, it is best to fix the base potential

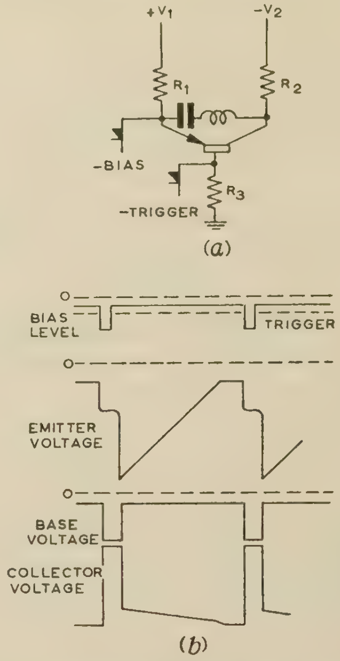


Fig. 3.—Pulse generator with base resistance.

(*a*) Circuit.
(*b*) Waveforms.

by a diode returned to a suitable negative supply. When the base is taken negative with respect to the emitter, the transistor turns on. Since the collector current exceeds the emitter current, the net base current drives the base negative; this improves the regenerative action and tends to sharpen the pulses. The potential of the emitter, which is conducting, follows that of the base, remaining slightly above it during the pulse. The voltage excursions at base and collector are proportional to R_3 and R_2, respectively, as is demonstrated in the Appendix.

Even if no negative output is required, the use of base resistance may be advantageous. It improves the regenerative action: it ensures definite triggering when the trigger pulse is much shorter than the output, because the immediate large shift in base and emitter potentials disconnects the triggering source. Usually less time and energy is needed to trigger on the base than on the emitter. An output from the base of one transistor is at a convenient level for triggering another. A disadvantage of using base resistance is that the triggering level depends on the standing collector current (which is a rather variable quantity) unless the quiescent base potential is fixed by a catching diode.

Any combination of positive and negative trigger and output may be used. Circuits to be described in Sections 3 and 4 use both triggers and both outputs at once.

(2.3) Frequency Division by a Small Factor

The circuits described above may be used for frequency division over a narrow range of driving frequency. To divide by a factor n, the given pulse train is applied to a trigger circuit whose recovery time is chosen to fall between $(n - 1)$ and n periods; having responded to the first pulse of the train, the transistor is biased off until after the $(n - 1)$th pulse but is free to respond to the nth.

With a circuit of the same configuration as that shown in Fig. 2(*a*), but with suitable component values for division by 3, the waveforms shown in Fig. 4 are obtained. To ensure operation from trigger pulses much shorter than the output pulse, either the impedance of the trigger source must be high (e.g. the

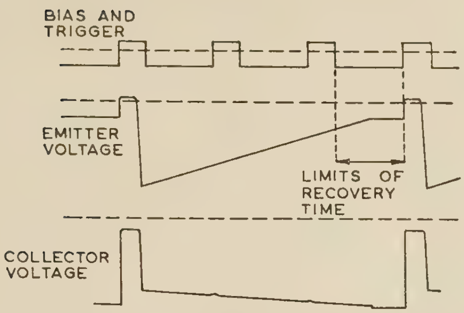


Fig. 4.—Waveforms of simple frequency divider.

collector of another transistor) or base resistance must be inserted so that the immediate change of potential disconnects the trigger source. This point is of some importance in a long chain of dividers, where it is often necessary or convenient to increase the pulse length as the repetition rate falls in successive stages.

An example of base coupling is shown in Fig. 5. This circuit

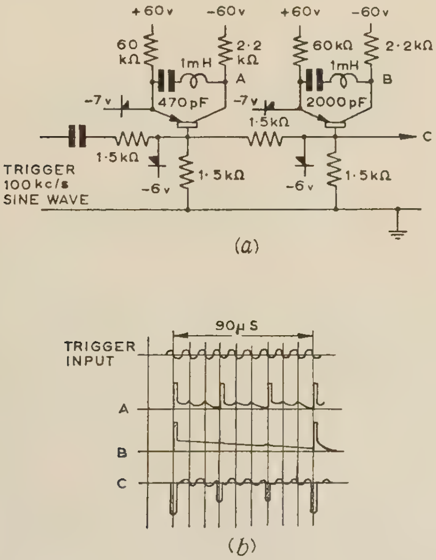


Fig. 5.—Two-stage frequency divider.

(*a*) Circuit.
(*b*) Waveforms.

is a 2-stage divider by 9, which operates from a pulse or sine-wave input at about 100 kc/s.

The division ratio obtainable depends on the extent of probable fluctuations in recovery time. Division by 2 or 3 is stable with normal components and power supplies. Division by 4 or 5 is stable if supply-voltage fluctuations are only a few per cent; larger changes can be tolerated if the positive and negative supplies vary together, for instance if they are derived by separate rectifiers from a common a.c. supply. For long-term reliability, stable components (especially C and R_l) are necessary, and an ageing transistor may give trouble sooner in a divider than in a simple pulse circuit. Division by a number greater than 6 is probably not worth attempting without stabilized supplies and catching diodes to standardize the waveforms.

(2.4) Frequency Division by a Large Factor

The type of circuit so far described is not suitable for division by large factors because the reset time varies rapidly with change of supply potentials or properties of devices. To achieve stable division by factors of 20–30 in one stage, a form of timing dependent only on passive components has been introduced.

The basic circuit has regenerative feedback through a pulse-forming circuit, shown in outline in Fig. 6, which may be a delay

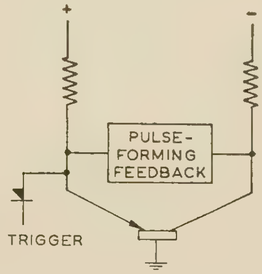


Fig. 6.—Schematic of basic pulse generator.

line [as shown in Fig. 1(*a*)]. It produces a single pulse when triggered by raising the emitter potential above that of the base. The action may be inhibited by a bias potential on the base: then, if a gating pulse is applied to the base, the transistor is triggered by whichever pulse in the timing train falls within the period of the gating pulse (Fig. 7).

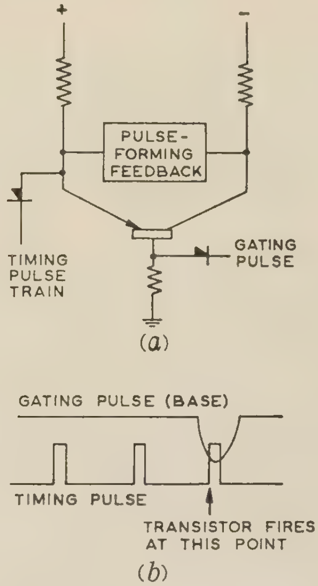


Fig. 7.—Gated pulse generator.

(*a*) Schematic.
(*b*) Gating waveforms.

The essence of the stable divider is that a gating pulse is obtained from the operating pulse by delayed reflections from a pulse-spacing network (some form of delay line) in the base circuit [Fig. 8(*a*)]. If the timing pulse spacing is t_0, it is obvious that division by a factor of n can be obtained by recirculation through a line of length nt_0; for the arrangement to be useful in practice the required line must not be prohibitively large or expensive and the tolerances on the line and other components must be reasonably wide. The length of line required can be reduced by using several traversals. An open-circuit line in the base lead of length $\frac{1}{2}nt_0$ returns a suitable gating pulse after one reflection from the far end [(i) of Fig. 8(*b*)]. A short-circuit line of length $\frac{1}{4}nt_0$ returns an inverted pulse (the wrong polarity for gating) after $\frac{1}{2}nt_0$; this is reflected without inversion from the transistor, which when non-conducting presents a high impedance to the line, and a gating pulse returns after a second reflection at the short-circuit [(ii) of Fig. 8(*b*)].

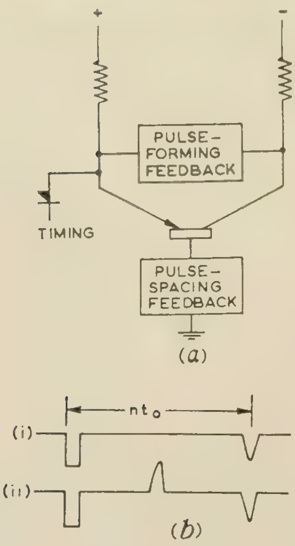


Fig. 8.—Self-gating frequency divider.

(*a*) Schematic.
(*b*) Self-gating waveforms.

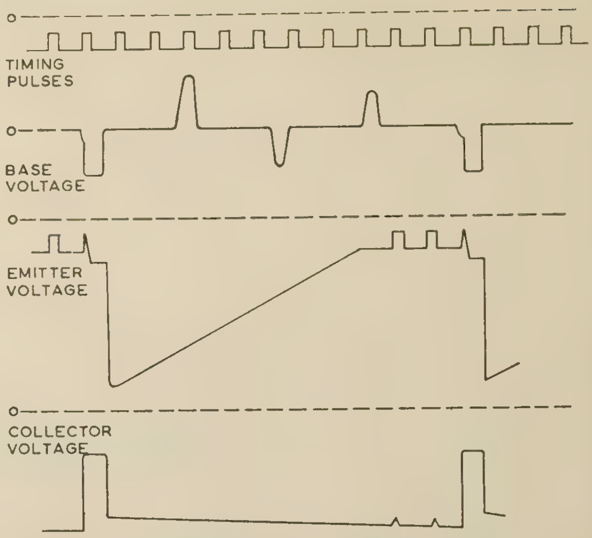


Fig. 9.—Waveforms showing division by a large factor.

Further reduction of the line length by a factor of 2 or 3 can be obtained by choice of recovery time. Simple dividing circuits are so arranged that recovery occurs after a time which exceeds $(n - 1)t_0$ but is less than nt_0, permitting the generator to respond to the nth timing pulse. This principle may be applied to the present circuit by using a short-circuit line of length $\frac{1}{4}nt_0/r$, and arranging recovery to occur after a time which exceeds $nt_0(r - 1)/r$ but is less than nt_0. The generator operates on coincidence of the nth timing pulse and the rth gating pulse. The stability of recovery time required is slightly greater than that suitable for division by factor r; this may be 2 or 3 with great reliability, and possibly more if the supplies are stabilized. This method is used in the time-division selector circuit described in Section 4.

The resolution required in the line depends on the pulse widths and amplitudes and on the permitted tolerances of bias potentials and delay time. As a rough guide, the rise-time of the reflected pulse is conveniently made about $\frac{1}{2}t_0$, so that the ratio between delay and rise-time is $2n : 1$; then the number of sections traversed is required to be about $4n$. If the line is traversed 8 times, the number of sections in the line is about $\frac{1}{2}n$.

An example in which a short-circuited line of length $\frac{1}{8}nt_0$ is used, with a factor of 2 gained by reliance on recovery time, is illustrated in Section 4. The waveforms for this arrangement are shown in Fig. 9.

Several other basically similar arrangements have been conceived which may turn out to be equally practicable or to have some advantages. There are two feedback functions (pulse-forming and pulse-spacing, respectively) and two positions where networks may be placed to obtain regeneration (emitter-collector admittance and base impedance, respectively). The allocations of functions to the positions could be interchanged, or networks for both functions placed in one or other position. The timing waveform, which has been shown as applied to the emitter, could equally well be applied to the base. One further example of a tried circuit must suffice (Fig. 10).

(2.5) Interlacing of Pulse Trains

A number of generators may be required to operate either in a permanent sequence or in an assignable sequence chosen from a repertoire: the first mode is characteristic of time-division transmission, the second of switching circuits. In either case, sequential operation implies that each generator is normally inhibited from response to timing pulses and is released by a specific signal: this may be accomplished by gating on the base [Fig. 7(*a*)], as in the divider described above.

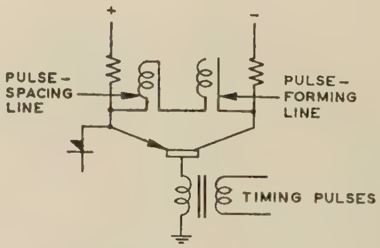


Fig. 10.—Alternative form of frequency divider.

For a generator with a permanent time position, the gating pulse can be derived by static delay from a generator earlier in the sequence. Also, in general, an output must be provided for controlling the following member of the sequence. This is described in Section 3.2.

The generator with a range of assignable time positions requires at least one pulse from an external source to define the position chosen, but thereafter generates its own gating pulse and so retains that position. The arrangement is similar to the dividing circuit considered in Section 2.4 and is described in Section 4.1.

(2.6) Transistor Operating Speed

There are three classes of circuits which are frequency-limited by transistor properties in different ways:

(*a*) Linear amplifiers and sine-wave oscillators, which depend on steady-state frequency response.
(*b*) Non-bottoming pulse circuits, in which, despite non-linear operation, the transition rates attainable are usually related to the steady-state frequency response.
(*c*) Bottoming trigger circuits, in which, mainly because of hole storage, transition rates are not related to steady-state properties.

The speed attainable in the circuits described has been found to depend on the small-signal steady-state frequency response of the transistor, the frequency response of the external circuit, the magnitude of the regeneration, and the magnitudes of the currents in the transistor.

Limitation by the external circuit is noticeable only if the time-constants of the feedback path are long in relation to those of the transistor. This can occur with the tuned-circuit feedback shown in Fig. 1(*b*), for which the relation quoted in Section 2.1 has been roughly verified. Usually it can be avoided by use of a fairly simple lumped delay line of a few sections: particularly simple networks are permissible if the load is not resistive but either is inductive or takes the form of a biased diode switch, so that little current is drawn initially.

With low currents, an adequate feedback network and only just enough regeneration, the rise-time is closely correlated with the steady-state properties of the transistor: the in-phase component of current gain at a high frequency, say 1·5 Mc/s, appears to be the best comparative measure, but the 3 dB α cut-off frequency is almost as good. Measured rise-times in these conditions range from 0·1 to 0·7 microsec for various specimens of point-contact transistor.

With greater regeneration, using a feedback admittance of several times the minimum value, faster and more uniform operation can be obtained. Typical circuits have given rise-times in the range 0·1–0·2 microsec, with collector current of about 100–120 mA and emitter current of 60–80 mA. To obtain maximum speed and power, a feedback transformer to augment the current gain is useful. Pulse power of 2 watts (40 volts and 50 mA swing) is readily obtained with this speed.

The time required to turn off a transistor can be much longer, owing to a hole-storage effect; when the transistor has been conducting heavily, collector current persists for a time after the emitter is biased off. The period varies considerably from one transistor to another, and with the magnitude and duration of the preceding emitter current. The best way to reduce and standardize the turn-off time is to maintain the collector potential sufficiently negative to absorb all the emitted holes, and this can be done by means of a catching diode returned to a suitable negative potential.[2,3] A collector-to-base potential difference of 4 volts is generally adequate. In many applications of these circuits, the dominant load has been some form of biased switch which conducts heavily when a threshold voltage is exceeded: in such a case, separate provision of a catching diode is not usually needed.

(2.7) Junction Transistors

The circuits described above have something in common with the blocking oscillator: they differ from it in that the combination of gain and phase available in a point-contact transistor makes the usual transformer unnecessary. The use of a transformer to augment the current gain increases the resemblance: it is then a short step further to a circuit in which all the current gain comes from the transformer, the transistor being a junction device with α less than unity.[4]

Without describing a variety of circuits in detail it may be stated that all the functions described have also been accomplished using junction-transistor blocking oscillators.*

(3) APPLICATION TO TELEPHONE TRANSMISSION

(3.1) Methods of Multiplex Pulse Generation

For the generation of precisely timed interlaced pulse trains there are three main possibilities:

(*a*) A master generator at the sampling rate, driving a tapped delay line, with a triggered generator driven from each tap.

(*b*) A master generator at the pulse repetition rate (i.e. the product of the sampling rate and the number of time positions) driving a binary counter, which produces the channel pulse at the outputs of a tree or matrix of diode switches.

(*c*) A master generator at the pulse repetition rate, driving a ring counter with one stage for each channel.

The choice between these methods has been guided by the properties of transistors and by experience with the simple pulse generators so far described. The main factors are:

(i) There is a maximum speed of operation. At present, due mainly to hole-storage effects, higher rates have been attained in generators of the type described than in conventional bistable circuits.*

(ii) There is no driving point of negligible impedance or admittance; an amplifier has less power gain, and a regenerative circuit requires more triggering power, than if hard valves were used.

(iii) The pulse generator described in Section 2 is a versatile basic circuit, with two trigger inputs and two outputs, and not too critically dependent on transistor characteristics.

The binary counter and diode-matrix system is most readily used with bistable circuits; the frequency limit is rather low in relation to the properties of the transistors. However, it offers the advantage of economy in components when very large numbers of channels are required, and may be made practicable by developments in transistors or circuits.

The delay-line method, which is widely used in hard-valve multiplex systems, cannot be applied to transistor systems in its simplest form because of the heavy loading of the line by low-impedance transistor circuits. It would probably be feasible if the pulse were regenerated at several points along the line; in this case, to avoid cumulative errors of timing, each pulse regenerator would have to be controlled from a master source. The arrangement would then be similar to the method which has been adopted; the latter could be considered as a delay-line system with pulse regeneration at every stage.

(3.2) The Time-Selective Ring Counter

The arrangement which has been found practically useful is a time-selective ring counter in which the basic pulse circuits are similar to those already described. The sequence of operation is controlled by couplings between the individual stages; the precise timing is fixed by a master pulse generator working at the overall pulse repetition rate.

Fig. 11 is a block schematic showing the functions required in the circuits; Fig. 11(*a*) gives more detail of the blocks shown as channel units in Fig. 11(*b*). Explanatory waveforms are shown in Fig. 12; they are given the same identifying letters as the associated points in Fig. 11.

Waveform A (Fig. 12) is a timing pulse train supplied to all channel units. Each channel unit contains a monostable pulse generator which is triggered by the coincidence of two signals: a timing pulse (A) and a delayed output from the previous channel generator (B). Each generator gives two outputs; the main channel output (C) and a delayed output (D) which is fed to the following channel. The pulse generator of the following channel is triggered by coincidence of this delayed output and the next timing pulse, and so on.

The basic circuit can provide all the functions of a channel unit if the two possible inputs and two outputs are fully employed. The positive timing pulse (A) is applied to the emitter; the negative sequence control pulse (B) to the base. Neither is large enough by itself to overcome the emitter-to-base bias, but the coincidence of the two triggers the transistor. The main output (C) is taken from the collector; the sequence-controlling output (D) comes from the base with the right polarity for connection to the base of the following channel unit. Fig. 13 shows

* When this work was started, junction transistors which could attain the speeds required were not available. In consequence, numerical values in the paper relate only to point-contact devices: the junction-transistor analogues have been realized only at lower speed and power. With the advent of improved junction transistors similar speed is now possible, although at somewhat lower power levels.

* The circuits described by Chaplin[1] which combine speed with bistability, were not published at the time this work was started, and in any case are unsuitable for providing the large pulse power required in this application.

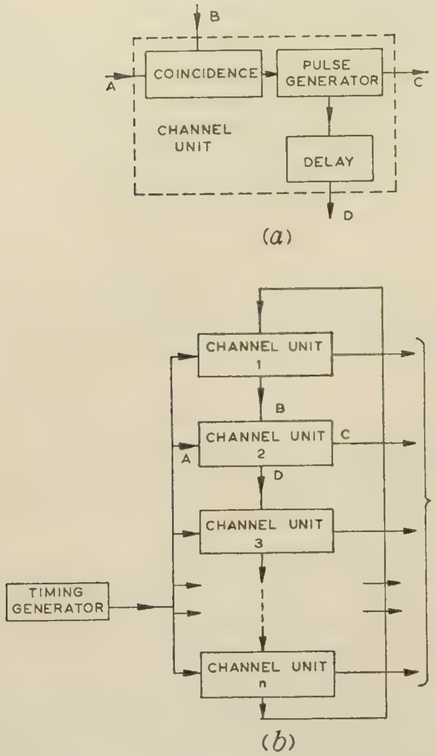


Fig. 11.—Schematic of multiplex pulse generator.

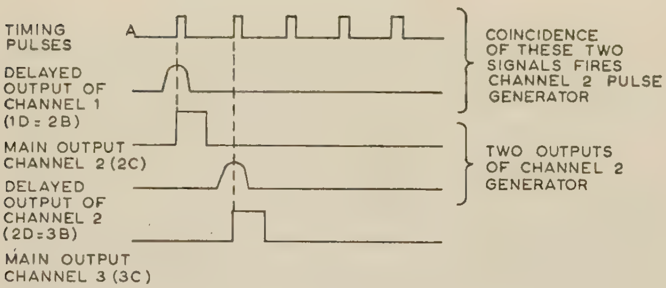


Fig. 12.—Waveforms illustrating pulse sequence control.

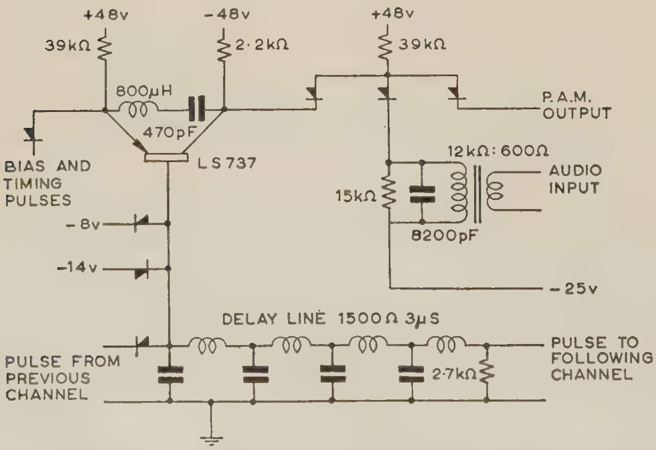


Fig. 13.—Unit of multiplex pulse generator.

a working circuit of one unit, with component values suitable for generating 2 microsec pulses at a 250 kc/s overall repetition rate; Fig. 14 shows the waveforms. The diode gate in the collector circuit is an elementary form of pulse-amplitude modulator.

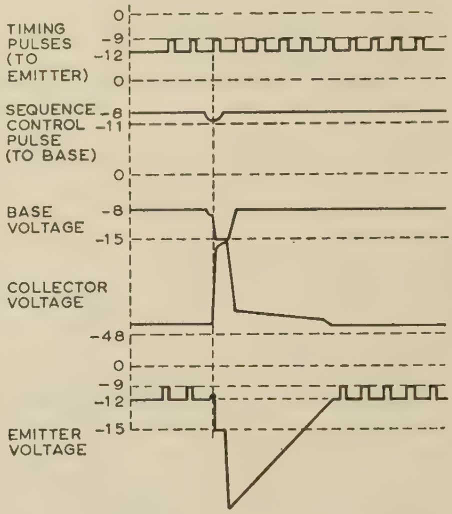


Fig. 14.—Waveforms of multiplex pulse generator.

(3.3) Transmission Terminal Equipment

The techniques described have been applied in the construction of two terminals providing 24 speech channels by time division. In later work, the elementary sampling gate shown in Fig. 13 has been abandoned in favour of a more efficient form of modulator which is described fully in two companion papers.[5,6] The form of the transmitted signal is then as shown in Fig. 15(*b*). The blocks representing the channel pulses indicate the range of excursion due to full modulation; both polarities are transmitted, and there are no carrier pulses in the absence of a channel signal. A complete frame occupies 108 microsec, of which 96 microsec is allotted to the 24 signal channels (2 microsec pulse and 2 microsec guard space); 4 microsec to a spare channel, which could be equipped for maintenance and supervision; and 8 microsec to the main synchronizing channel (6 microsec pulse and 2 microsec guard space). The base level between channel pulses is modified in alternate spaces by interpolating a small positive 2 microsec pulse, conveying a fine synchronizing signal. This is suppressed when it would otherwise coincide with the 6 microsec pulse.

A block diagram of one terminal, designated A, is shown in Fig. 15(*a*). Terminal A contains the master timing source, a 250 kc/s crystal oscillator, and transmits synchronizing information. Terminal B is inert until synchronizing signals are received, upon which it starts up in the correct phase. The channel units and the main synchronizing generator of terminal A are connected in a ring, as described in Section 5, for normal working; the time-base of a cathode-ray-tube monitor is driven at the sampling frequency from a unit in the ring. Should a unit fail, the count around the ring stops. By pressing the test key, the monitor and the synchronizing channel can be triggered at the sampling rate from the output of a frequency divider while a d.c. signal is applied to all channels. Each unit fires in sequence after the synchronizing channel until the fault is reached: it is then apparent from the monitor display which channel unit has failed. All channel-unit faults so far encountered have resulted either in complete failure, detectable as described, or in free running of a pulse generator at an uncontrolled frequency; the

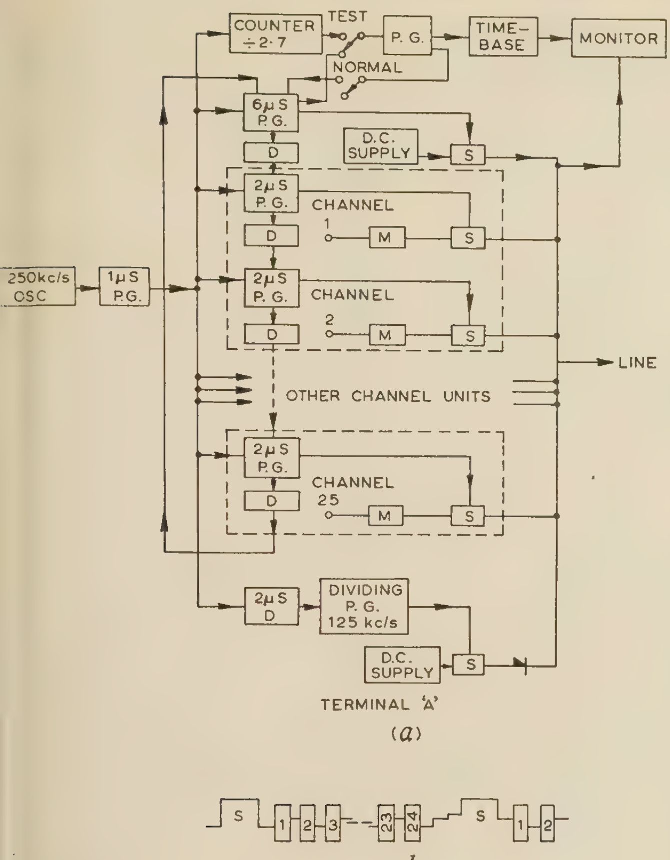


Fig. 15.—Transmission equipment.

P.G. = Pulse generator. D = Delay.
M = Modem. S = Switch.
(*a*) Schematic.
(*b*) Waveform.

atter condition is less easily traced, but its occurrence is obvious rom the monitor display.

All the pulse generators and frequency dividers are of the ypes described in Sections 2.2, 2.3 and 3.2. The terminal B is imilar, except that timing is derived from the incoming synchronizing signals. This apparatus has operated for periods of several thousand hours, and at moderate ambient temperature is quite reliable. Rise of temperature is the worse enemy: it may cause the standing current of a transistor to rise, thereby releasing the base from its clamped potential (8 volts in Fig. 13) and allowing the circuit to run freely. In later work (such as hat described in the next Section), the resistance in the base circuit has been reduced or eliminated.

(4) APPLICATION TO TELEPHONE SWITCHING

(4.1) A Time-Division Selector Switch

It has for some time been recognized that multiplex techniques nay be used, not only to make manifold permanent connections

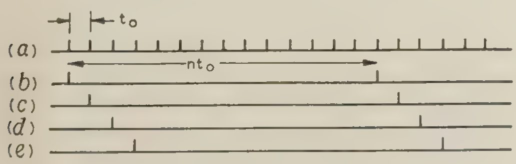


Fig. 16.—Interlacing of pulse trains.

as in transmission systems, but also to make controlled temporary connections as in switching. Since one time position is distinguishable from another, not *per se* but only by some external reference, it is possible to make one unit capable of operating on any channel.

Such a unit can be built around the single-stage frequency divider described in Section 2.4. A dividing circuit of the type described, if fed with timing pulses spaced at t_0 [Fig. 16(*a*)], can operate repetitively at intervals nt_0 in any one of n time positions [as Fig. 16(*b*)–(*e*)]. The circuit is stable in the 'off' position; if it is triggered initially at any given time position, it continues to run at that time position until it is stopped. Thus, it can be used to close an electronic switch between a line circuit and the busbar of a time-division multiplex system, thereby putting the line circuit in connection with any chosen channel of the multiplex signal. The essential features are, first, division by a large factor n (20–30) in one stage, secondly, the retention of a time position specified initially by a single pulse.

A complete selector unit is shown in Fig. 17(*a*), and its waveform in Fig. 17(*b*). The unit is designed to work on a 25-channel basis with 10 kc/s sampling. The fourth reflected pulse from a 12·5 microsec delay line is used to gate the 250 kc/s timing pulses. The electronic switch is in the form of a reciprocal pulse modem,[5,6] which, through storage of energy in a reactive network (delay line and low-pass filter), converts a continuous signal into a modulated pulse train, or vice versa, with very little power loss. The switch is held off by a static bias of 15 volts when the pulse generator is not running; this is enough to prevent pulses from other units connected to the same busbar from breaking through. When the pulse generator is running, dynamic bias of about 30 volts is available to isolate the given channel circuit from other signals on the busbar. The voltages and currents of the switching circuit may require some reconsideration to obtain the best combination of linearity, crosstalk margin and peak power; the peak pulse power which can be utilized in the present circuit is about 10 volts into 400 ohms, namely 250 mW, corresponding to about 4 mW peak audio power in the receiver output (2 mW r.m.s. on sine-wave signals).

The construction of the delay lines is not shown in the diagram. A design due to H. Feissel[7] has been found useful: the 12·5 microsec line used in the experiments had 16 sections.

As noted in Section 2.4, several arrangements are possible; both that shown and a slightly different circuit with timing pulses applied to the base have been thoroughly tested for reliability of starting from a single trigger pulse in the correct phase. About 2 500 operations were performed correctly in a repetitive testing arrangement, which is enough to inspire a modest confidence in the device.

(4.2) Discussion of Properties

A proposed time-division switching scheme has to face two competitors. First, the various electro-mechanical switching methods, including not only the thoroughly entrenched systems but also new ones which, by using electronic control of mechanical switches, have shown that connection and control are separate problems, and that the use of electronic devices in each requires separate justification. Secondly, various existing or proposed devices for effecting voice-frequency connections, of which at least one (the gas gap) has reached a very promising state of development.

An electronic switch to compete with an electro-mechanical switch for speech paths should have some of the attributes of a normal rotary selector, namely:

(*a*) It should transmit in both directions.
(*b*) It should have a low loss.
(*c*) It should be self-holding, requiring little or no central storage capacity, unless this can be provided very simply.

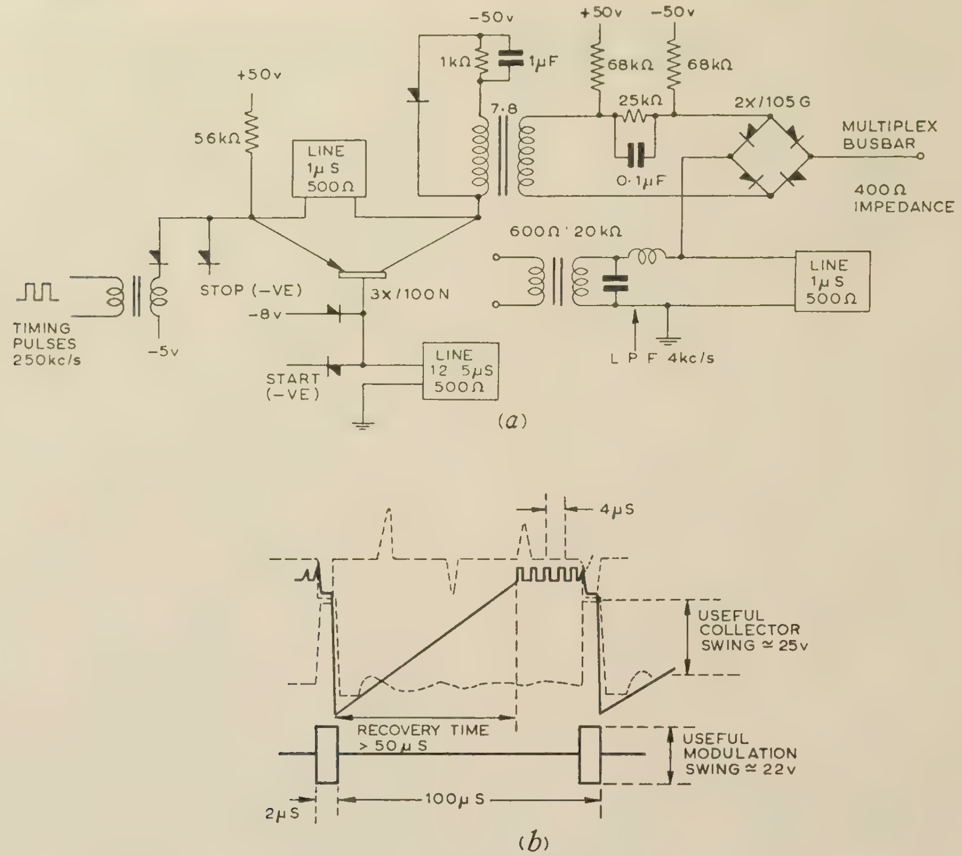


Fig. 17.—Time-division selector switch.
(*a*) Circuit. (*b*) Waveforms.

(*d*) A unit of moderate size and cost should correspond to a whole selector, not just a single pair of contacts.

In general, voice-frequency electronic switches fail on condition (*d*), and possibly also on (*c*). Multiplex switches, which inherently have a good chance of meeting condition (*d*), have in the forms previously proposed failed on conditions (*a*), (*b*) and (*c*). A selector of the form described here has all four properties. The reciprocal pulse modem fulfils (*a*) and (*b*); the frequency-dividing pulse generator fulfils (*c*); while point (*d*), as noted before, is inherent in time-division methods and is bound to be fulfilled if, as in this case, the apparatus is fairly simple.

This selector is, in fact, a good electronic analogue of a uniselector. It transmits in both directions with a low loss, and, having been triggered at a given time position, it operates repetitively at that position until it is stopped. A multiplex selector controlled by this means is like a rotary switch, which, having been turned to a given spatial position, remains there until it is released. It is therefore a useful building brick for an electronic telephone-switching system.

It is worth noting that the general idea is not tied to the use of transistors. One active device per selector is required; this could easily be a hard valve, and possibly a thyratron.

(4.3) Time-Division Storage

The time-division selector incorporates a store, since the identity of one channel from a group of 20–30 is retained. Static storage would require at least 5 binary devices, with some means of coupling them to the speech path.

It is interesting to assess the merits of this method of using the delay line, as compared with storing as many binary digit as possible by recirculation. The ratio of delay to rise-tim advocated for the present circuit is about n : 4, for n channels this allows a good margin of safety and could possibly be reduced Assuming that pulses spaced by a period equal to the rise-tim can just be separated, the line could store $\frac{1}{4}n$ binary digits Comparing this with the storage of one of n time position (i.e. $\log_2 n$ binary digits), it is seen that there is little differenc for $n = 12$ to $n = 24$. The line embodied in the time-divisio selector is therefore being used efficiently, and use of simila circuits in a register to retain decimal digits would seem reasonable.

(5) ACKNOWLEDGMENTS

The writer is indebted to his colleagues Mr. J. C. Price an Mr. R. B. Herman for contributions to this development, an to Standard Telecommunication Laboratories Ltd. for pe mission to publish the paper.

(6) REFERENCES

(1) Chaplin, G. B. B.: 'A Point-Contact Transistor Scalin Circuit with 0·4 microsec Resolution', *Proceedings I.E.E* Paper No. 2064 R, March, 1956 (**103 B**, p. 505).

(2) Cooke-Yarborough, E. H.: Contributions to Discussio *Proceedings I.E.E.*, 1953, **100**, Part III, p. 245.

(3) Price, J. C.: 'A Conductivity Storage Transistor Puls Width Modulator', *Electronic Engineering*, February, 195 **30**, p, 88.

(4) Linvill, J. G., and Mattson, R. H.: 'Junction Transist Blocking Oscillators', *Proceedings of the Institute of Rad Engineers*, 1955, **43**, p. 1632.

(5) CATTERMOLE, K. W.: 'Efficiency and Reciprocity in Pulse-Amplitude Modulation: Part 1—Principles' (see page 449).
(6) PRICE, J. C.: 'Efficiency and Reciprocity in Pulse-Amplitude Modulation: Part 2—Testing and Applications' (see page 463).
(7) FEISSEL, H.: 'Lignes à retard à constantes localisées', *L'Onde Éléctrique*, 1954, **34**, p. 53.

APPENDIX: IMPEDANCE RELATIONS IN THE PULSE GENERATOR

It is not proposed to enter into a full analysis but to give a simple treatment from which the essential design relations emerge. We assume that regeneration occurs, and consider currents and impedances in the bottomed or near-bottomed condition during the pulse. These are represented in Fig. 18. The

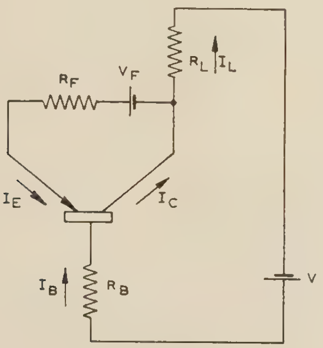


Fig. 18.—Impedances and currents during a pulse.

use of resistance notation does not imply limitation to physical resistances: each value specifies only the ratio of potential difference to current during the pulse. Thus, in Fig. 17, R_B and R_F are delay lines, whose characteristic impedances are the relevant values: R_L is a diode switch, whose ratio of voltage swing to peak current (referred to the transformer primary) gives the value of R_L. It is assumed that internal r_e and $1/r_c$ are small in relation to R_F and $1/R_L$, respectively, which is usually correct: internal r_b may not be negligible but can be added to R_B. The potential differences between all electrodes are assumed negligible: catching action to prevent bottoming, if present, is readily allowed for. The electromotive force V_F in the feedback path is the p.d. to which the feedback line or capacitor is charged immediately before the pulse: this is normally a large fraction of the supply potential V.

The basic relations are:

$$I_E = V_F/R_F$$
$$I_L R_L + I_B R_B = V$$
$$I_C = I_L + I_E = I_B + I_E$$

from which we can deduce the voltages and currents in a given circuit:

$$I_L = I_B = \frac{V}{R_L + R_B}$$

$$\text{Pulse amplitude at collector} = \frac{VR_L}{R_L + R_B}$$

$$\text{Pulse amplitude at base} = \frac{VR_B}{R_L + R_B}$$

The current gain in the working condition is

$$\bar{\alpha} = 1 + \frac{I_B}{I_E} = 1 + \frac{V}{V_F}\frac{R_F}{R_L + R_B}$$

and for regeneration this must be substantially less than the linear current gain α. Assuming that (as is usually the case) $V_F/V \simeq 1$ and $\alpha \simeq 2$, a reasonable condition is then

$$R_F < R_L + R_B$$

As an example of heavier regeneration, in selector circuits as shown in Fig. 17, $V_F/V \simeq 0{\cdot}8$ and $\bar{\alpha} \simeq 1{\cdot}4$.

Practically, one will usually know the required values of pulse amplitude at collector and base, decided with reference to the function of the generator and the voltage limits of the transistor: the likely value of V_F/V, set by the bias and triggering arrangements: and a target value of $\bar{\alpha}$, depending on the compromise between transition speed and hole storage but usually in the range $1{\cdot}2$–$2{\cdot}0$. The ratio R_B/R_L follows from the ratio of pulse amplitude at base and collector: the value R_L, from the load current or power. The feedback impedance is then

$$R_F = (\bar{\alpha} - 1)\frac{V_F}{V}(R_L + R_B)$$

This defines the impedance of a feedback line, where one is used: in the case of tuned-circuit feedback, to which this analysis is only a rough approximation, it can be taken as a trial value of $\sqrt{(L/C)}$.

DISCUSSION ON THE ABOVE THREE PAPERS BEFORE THE RADIO AND TELECOMMUNICATION SECTION, 19TH MARCH, 1958

Mr. T. H. Flowers: The transmission system using the pulse modems described in these papers was invented in three different places more or less simultaneously—by Haard and Svala in Sweden, by one of the present authors, and also by French of the Post Office. I agree with the authors that the chief use of this form of transmission is likely to be in electronic exchanges, and all my remarks refer to this application.

The paper on the design of pulse modems is by far the most complete and satisfactory treatment we have yet had of this subject. Two questions seem to be important. First, the analysis and testing assume a resistance termination of the modems, but in practice the terminations will be transmission lines, which may often have an impedance which varies with frequency and is very reactive. What effect will this have on the transmission characteristics? In particular, I would think that anything as refined as a 12-element network would be unjustifiable with an uncertain termination, and this is probably the reason why the authors think that the 4-element network is sufficient.

The second point of interest is what the authors call the transparency of the modems. With a 2-wire switch system, terminal amplifiers on junction circuits must have compromise balances, and it is difficult to find a compromise balance now. It might be rather worse when viewing line impedance through pulse modems, and if this is true, line amplification will be more difficult. This will necessitate low-loss transmission (2 dB or less) through the modems, i.e. the same order as present electromechanical switches.

On testing, electronic switches insert within an exchange a transmission link having characteristics similar to those of a transmission circuit between the exchanges. We specify trans-

mission circuits between the exchanges in respect of the frequency pass-band, crosstalk (usually specified to be a minimum of 70 dB, which is much more than is quoted in the third paper), harmonic and inter-modulation products, overload capacity, etc. A similar, but perhaps rather more severe, specification should apply to transmission within exchanges, and testing according to this specification should suffice. However, the papers describe effects such as 'buzz', 'burr' and 'mush', which require subjective tests with a number of people using actual speech. I do not consider their results to be conclusive. For the number of tests performed the 95% confidence limits are rather wide. In addition, one must remember that in a long connection there may be not one but eight or ten, stages of switching, with transmission lines between. This does not mean that the defects will be multiplied by the same factor, since some of the distortion products will be attenuated in the network. But I am sure that more testing is required before we are certain that the transmission would be satisfactory in a national system.

The papers describe some applications of this method of transmission to switching systems, including a v.f. dialling system which could be applied to any telephone system, whether it uses reciprocal gating or not. For the rest of the applications, only time will decide their value. Certainly the selector circuit shown in Fig. 17 of the third paper must be the simplest electronic connector switch which has so far been disclosed. I am not quite clear how the author proposes to use it. If it is on a basis of one per subscriber's line, even though it is simple it would be rather extravagant overall. If it is not on this basis, I do not know how it could be fitted into the system. I should like more information on this point.

The advantages of the pulse modem method of switching, such as simplicity and power economy, are adequately stated, and I have mentioned some of the difficulties. Competitive systems, such as 4-wire multiplex and space-switching, similarly have advantages and disadvantages. It will be interesting to observe in the next year or two the progress made by these various methods.

Mr. L. I. Farren: Electronic switching techniques, which have introduced a whole range of new electronic devices and methods, are at present slightly suspect to transmission engineers. Until recently, the transmission characteristics of exchanges have not been an important factor in the overall transmission plan. Now they have become vitally important, and in the electronic exchange the factors of crosstalk, intermodulation, frequency response, etc., will all have to be considered. For example, there might appear to be an immediate advantage in simplicity in using a 2-wire system as proposed; at the same time, degradation in certain transmission factors by the use of such a system might outweigh the economic advantage.

The author describes a 25-channel p.a.m. system, and one might infer from the paper that multiplex units of more than 25 channels would be difficult to realize. If one considers a large exchange of, perhaps, several thousand lines, it will be necessary to connect numbers of these 25-channel units together. I should like to know how many might be connected together without significant degradation in transmission performance.

An example is quoted of 46–66 dB crosstalk for a system employing point-contact diodes and 55–75 dB with junction transistors. This compares unfavourably with other types of t.d.m. systems operating on a 4-wire basis, in which no difficulty is experienced in achieving 75 dB crosstalk in a 100-channel system. These 4-wire systems are at first sight less economic, but they have the merit of technical simplicity.

One practical point caught my attention. The impedance of the low-pass filter depends on the ratio of the repetition sampling duration to the pulse duration. In the example described this ratio is 50 : 1 and is equal to the ratio between the filter impedance and the delay-line impedance. With a reasonable delay-line impedance (namely 500 ohms) we have a filter of 25 kilohms impedance, which is rather high. The inductance in such a filter will be large and might be difficult to realize. To obtain more than 25 channels the pulse duration must be reduced, the ratio of repetition duration to pulse duration increased, and consequently the low-pass-filter impedance increased, with a consequent increase in inductance (unless the delay-line impedance is decreased at the same time). What is the maximum number of channels which might be achieved subject to this practical limitation?

Mr. G. H. Parks: We have done some experiments using a symmetrical transistor as the switch to close the path between the two storage devices. Briefly, using a 10 kc/s repetition rate and a 4-element filter followed by a tuned circuit, an overall loss of slightly less than 1·5 dB has been achieved through two switches. This loss is obtainable up to a pulse-duty ratio approaching 50 : 1, but for higher ratios the loss increases, and it is disappointing not to find in the papers any reference to the factors which limit the maximum number of channels which can be used.

The use of a symmetrical transistor here is rather interesting for, apart from a potential applied to the base to turn it on or off, no other supplies are needed. The speech power serves as its own supply, and there is no need even for a d.c. connection between emitter and base.

It is clear that this system suffers from certain imperfections in transmission which are not shared by t.d.m. systems in general. Incidentally I noticed an appreciable change in the quality of Mr. Cattermole's voice, as heard over the sound-reinforcement system, when the modem was switched out, but I think that this was due only to the restoration of the high-frequency response which was lacking when the modem was in circuit.

Finally, a comment on the paper on transistor pulse generators: such is the rapid progress in the transistor art that, although the point-contact transistor was only invented in 1948, a paper which describes circuits using point-contact, as opposed to junction, transistors is now something of a rarity.

Mr. R. B. Herman: The current/voltage characteristic for an ideal rectifying junction is theoretically $I = I_s(\epsilon^{eV/kt} - 1)$ which, for a forward current of 100 mA, corresponds to an incremental resistance of 0·265 ohm at 300° K. Diodes approaching this performance with sufficiently low capacitance and minority-carrier storage for operation at high frequencies are not yet available, but recently developed high-frequency symmetrical junction transistors provide an alternative to diode switches in the pulse modems.

The main difficulty in using present types of high frequency junction transistor in the pulse-generator circuits is their low voltage ratings. This is particularly important in the design of the circuits for frequency division by a large factor, which utilize multiple traversals of a pulse through a pulse spacing network. Because of attenuation in the network, an initial pulse amplitude of at least a few volts is usually required for reliable triggering. This sets a lower limit to the amplitude of the sawtooth recovery waveform across the pulse-forming network which adds to the p.d. between the emitter and base of the transistor. Reducing the attenuation in the pulse spacing network to a negligible value is, however, unsatisfactory, since apart from cost and space considerations, this results in an accumulation of energy in the network during successive operating cycles and hence to errors in the spacing of the output pulses. It is hoped that manufacturers will provide transistors with higher voltage ratings, having comparable ratings for the emitter and collector junctions.

In the pulse-generator circuits illustrated the leading edges of the output pulses are controlled by an external timing waveform, whereas the pulse durations are defined by separate pulse-forming networks in each circuit, which must be constructed to fairly close tolerances. An alternative arrangement is to use the external waveform, or two separate waveforms, to control both the leading and trailing edges of the output pulses, e.g. a rectangular waveform for which the positive part of each cycle has a duration equal to the required pulse width. If this is coupled to the emitter of a p–n–p transistor in a pulse-generating circuit, the leading and trailing edges of the output pulses may be controlled respectively by the positive- and negative-going transitions of the waveform. This technique gives better precision in the timing of the pulses and often simplifies the circuits, but generally requires a timing waveform generator with greater available power.

Mr. E. H. Cooke-Yarborough: The authors seem to have made good use in their circuits of the special properties of point-contact transistors. For several years we have had 324 such transistors in continuous operation. The failure rate has fallen steadily and now amounts to 0·2% per 1 000 hours. Most of our transistors are at least three years old, and I should be interested to hear what improvements in point-contact transistors have been achieved since these days and what reliability the authors now achieve.

Have the authors considered, as an alternative to the point-contact transistor, the use of the hybrid transistor described by Salow and von Münch,* which seems to have certain advantages?

Mr. W. E. Thomson: I should like to raise two points about the storage networks described in Paper No. 2474 R. The first suggests an alternative approach to the design of these. The second suggests an alternative ideal form.

As Mr. Cattermole has shown for band-limited systems, the ideal reactance network which controls the flow of energy between the storage capacitor and the source (or load) resistor, is an ideal low-pass filter. The paper then considers practical filters with Butterworth or Chebyshev characteristics. Modern insertion-loss design methods, however, suggest that the most efficient configuration is a series of m-derived π-sections. A single section, with three capacitors and one inductor, would probably be satisfactory, but the best values for the four design parameters have still to be found. It seems reasonable to make the frequency of infinite loss coincide with the switching frequency, thus helping to remove disharmonious components from the output.

There is another ideal form of reactance network which has certain attractions for the transmission of waveforms. With this, the storage capacitor, of value $\frac{1}{2}t_1/R$, completely charges or discharges linearly during the sampling interval; there is thus no distortion associated with the switching. The output waveform gives a rectangular approximation to the input waveform, i.e. between sampling instants the output waveform is constant and equal to the average value of the input waveform during the previous sampling interval. From the frequency-response aspect, however, the system is not satisfactory: we have

$$|H(j\omega)| = \frac{\sin \frac{1}{2}\omega t_1}{\frac{1}{2}\omega t_1}$$

so that the loss gradually increases with frequency, being about 4 dB down at half the switching frequency.

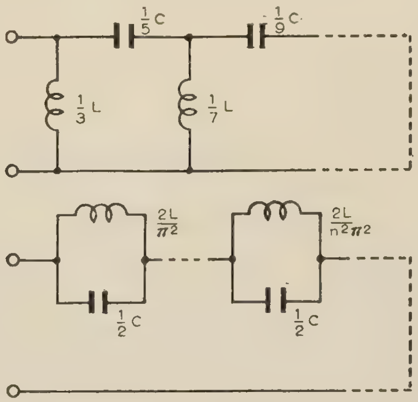


Fig. A.—Alternative networks to realize $Z(p) = R(\coth \frac{1}{2}pt_1 - 1/\frac{1}{2}pt_1)$; $L = \frac{1}{2}t_1R, \quad C = \frac{1}{2}t_1/R$.

These ideal characteristics require an infinite network for exact realization. It is of π-configuration, the shunt arms being the source (or load) resistor, R, and the storage capacitor ($C = \frac{1}{2}t_1/R$). The series arm is purely reactive, with an impedance $Z(p)$ shown in Fig. A. Practical networks are derived by using the first few elements of either infinite chain.

THE AUTHORS' REPLY TO THE ABOVE DISCUSSION

Messrs. K. W. Cattermole and **J. C. Price** (*in reply*): We learn with interest of yet another independent invention of the pulse modem, and now feel that the time is ripe for its exploitation.

Several comments and inquiries can be answered in the light of more recent work by ourselves or our colleagues. Endeavouring to be candid in describing the speech transmission qualities of the modems, we have perhaps given the onomatopoeic criticisms more prominence than they deserve. Even in the original tests many subjects either did not detect these phenomena or considered them present in only a trivial degree. However, we have established that much of the trouble was due to overloading, and that if this is avoided the only effect readily perceptible is bandwidth limitation. Mr. Parks's comment on the demonstration supports this contention. The original results are of value in showing that, despite perceptible overloading, intelligibility is still good.

The original work, as Mr. Flowers remarks, paid little attention either to the operation of modems between complex impedances or to the impedance presented by a modem to the line. Our colleague Mr. J. C. Emerson has since designed an improved network which approximates a match over a greater fraction of the pass band. Modems using this network approach very closely both the transmission characteristic and the impedance loci of a passive filter: thus there is no reason to suppose that modems will introduce matching or balancing problems of a new order of difficulty.

Point-contact-transistor circuits were described because, at the time of submission of the papers, suitable junction transistors were not available. Related circuits along the lines suggested in Section 2.7 of the pulse-generator paper, using the same general principles of timing and interlacing, have since been built; despite certain disadvantages we would now prefer to use them, because of the better availability and consistency of the junction devices. As noted by Mr. Herman, there is no junction transistor combining high speed with high voltage ratings on emitter and collector—a lacuna regretted by designers of pulse circuits.

The number of channels will certainly be limited by the impedance ratio attainable, as Mr. Farren points out; 25 kilohms,

* SALOW, VON H., and VON MÜNCH, W.: 'Über einen Schalttransistor mit kurzen Schaltzeiten', *Zeitschrift für angewandte Physik*, 1956, **8**, p. 114.

as used in our early work, is not unreasonably high for the audio circuit. The 500-ohm pulse-circuit impedance was chosen to suit the original electronic switches, namely point-contact diodes; later work has shown that 100 ohms is quite suitable for use with symmetrical junction transistors. Hence an increase to 125 channels is quite compatible with convenient magnitude of impedance. Reduction of pulse-path impedance with the same number of channels gives a proportionate improvement in crosstalk margins. Since the crosstalk performance was the least satisfactory property of the original circuits, we are pleased to record that, in later experiments, margins of 75–85 dB have been maintained even in the presence of substantial highway capacitance.

Leaving the solid ground of further study, we can comment only briefly on other points. The time-division selector could be used per subscriber's line in a small switchboard or concentrator, and we agree with Mr. Flowers that it would be extravagent in a large exchange, which would need substantially more than one per line. Our subjective tests were admittedly modest in scope, for which reason we presented them without the full-dress finery of statistical mathematics. Their only purpose was to establish that any extraordinary effects not detectable by normal transmission measurements were not unduly obtrusive—a conclusion which we think most of the audience at our demonstration would admit. The hybrid transistors cited by Mr. Cooke-Yarborough would appear to be suitable for our pulse generators: so would *p–n–p–n* or other thyratron-like structures or combinations of complementary transistors, but we cannot claim experience with any of these devices. Mr. Thomson's alternative storage network leads, as he says, to a poor amplitude/frequency characteristic, but its phase characteristic is good and might be of use in a system intended for digital data rather than speech. This network is also interesting in that it contains two parts determining delays of the pulse duration and pulse-interval respectively.

DISCUSSION ON 'SERVO-OPERATED RECORDING INSTRUMENTS'*

SOUTH MIDLAND RADIO AND MEASUREMENT GROUP AT BIRMINGHAM, 28TH APRIL, 1958

Mr. J. S. Roebuck: I understand that the strobing potentiometer was used when one had a fast wave at high repetition rate. It is often very useful to obtain a record of a waveform which cannot easily be taken on an oscilloscope and photographed afterwards. Is the method described satisfactory for obtaining a record without going to the trouble of photographing an oscilloscope, e.g. for the display of a square pulse at television repetition rate?

Mr. G. A. Montgomerie: Perhaps the author could make a few comments on the question of reliability. I am thinking, in particular, of those people who are afraid of radio valves and regard them as 'glass bottles'. I have no doubt that work is going on so that thermionic valves can be replaced by transistors in this sort of equipment.

The second suspect is the mechanical 'chopper'. Is this here to stay, or is a substitute on the way?

Dr. D. Karo: It is a pity that the author gives no details of a very important field of recorders which have been produced by several firms in France and other Continental countries. Some of these recorders act before the error occurs, such as when controlling magnetic flux. For example, before any change of flux occurs there is a change in H, and therefore the instrument works on the pre-feel before the error occurs, since there is a time lag between H and B. This is quite an important field in the future.

What is the reliability of instruments when the record is fluctuating, such as the load on rolling-mills? Has the author any experience of the reliability of recorders which record, say, the current or power in a case like that of a rolling-mill, where in a short time the power can vary and the frequency of variation will be of the order of that of the instrument itself, say 2–3 c/s? An ordinary ammeter reading the current of an induction motor driving a rolling-mill will not give a true reading, because there is fluctuation in the load.

It would be a good idea to draw the attention of users and manufacturers of controllers to the fact that there is an optimum damping in which the logarithmic decrement is exactly π, and where the sensitivity is 23% more than in critical conditions, the overshoot is very small.

Mr. J. B. Borthwick: Has the author any experience of the memory type of recorder? This more or less retains in its memory for a short interval an imprint of circuit condition.

A permanent record of any specific event is available should it be required. I am interested in this for the recording of high-frequency phenomena.

Dr. A. J. Maddock (*in reply*): One of the main features of the strobing-potentiometer recorder is that a permanent record is obtained, without photographing or developing, of a repetitive waveform. The strobe pulse is of 3 millimicrosec duration, so very rapid changes in waveform can be studied. The basic repetition rate of the pulse is a maximum of 15 kc/s, so that the waveform being studied must come within this range or be such that a sub-harmonic can be arranged to do so. Television pulses if repetitive, at the line-time-base frequency could therefore be recorded.

Considerable advances have been made in the reliability of thermionic valves, and special types are readily available which are no less reliable than some other components. Transistors and valves are complementary and both will be used in electronic apparatus. The mechanical 'chopper' continues to give satisfactory service, and some manufacturers have changed from other methods of inversion to use this device. Development is not sufficiently advanced to say whether transistors will be a serious rival, particularly for low-level signals.

Dr. Karo and Mr. Borthwick take the discussion outside the scope of the review, which is concerned solely with recorders which are servo-operated, and deals with the instruments themselves and not with applications. Both types of recorder mentioned by them are interesting and have potential uses. Commercial models are becoming available.

A servo-operated recorder cannot be expected to deal adequately with rapid fluctuations in input signal which are of duration similar to the speed of response of the servo system. Other types of recorder should be used in such cases.

* MADDOCK, A. J.: Paper No. 2131, September, 1956 (see 103 B, p. 617).

.1.315.23: 621.317.333.4

The Institution of Electrical Engineers
Paper No. 2710 R
Sept. 1958

THE FACTORS INFLUENCING THE USE OF PNEUMATIC METHODS FOR THE LOCATION OF SHEATH FAULTS IN PRESSURIZED TELEPHONE CABLES

By E. J. HOOKER, M.A., Ph.D.(Eng.).

(*The paper was first received* 13*th March, and in revised form* 14*th May*, 1958.)

SUMMARY

An expression, based on the classical Poiseuille equation for gas flow through capillary tubes, has been obtained to describe the passage of gas through non-coaxial telephone cables under steady-state conditions, and, for cables in this category, the factor governing the flow has been shown to depend upon the number, gauge and arrangement of the conductors. An approximation of the pressure gradient associated with a small rate of gas flow has been derived from this expression, and consideration has been given to the degree to which this limits the successful location of sheath faults by methods involving the direct establishment of the pressure gradient, the installation of pressure-sensitive contactors, rate-of-flow measurements and the use of flow-direction indicators.

LIST OF SYMBOLS

p_1, p_2 = Gas pressures at two positions in the tube or cable.
p_m = Mean cable pressure.
p_A = Atmospheric pressure.
p_c = Input pressure to cable.
p_l = Cable pressure at leak.
Δp = Maximum difference in contactor operating pressures.
δp = Actual difference between nominally identical pressures.
pv, pv_A, pv_B = Gas flow per unit time, for volume v, v_A, v_B, measured at pressure p.
pv_T = Gas leak per unit time, for volume v_T measured at pressure p.
dp/dl = Pressure gradient.
l = Capillary-tube length.
L = Cable length.
l_A = Distance between leak and end A of cable.
Δl = Distance between actual and estimated positions of leak.
l_c = Distance between contactors.
a = Capillary radius.
D = Conductor diameter.
n = Number of capillary paths in cable.
N = Number of conductors in cable.
$k = l/L$.
K = 'Make-up' constant for given cable type.
R = 'Pneumatic resistance' of unit length of cable.
S = Pneumatic constant for leak.
η = Viscosity of gas.

(1) INTRODUCTION

The problem of maintaining telephone circuits in working condition, even though the sheaths of the cables carrying them may be damaged and no longer capable of preventing the ingress of moisture, has been at least partly solved by keeping the interiors of the cables under a positive pressure of dry air, nitrogen or other inert gas. Several accounts of the use of such systems have already been published.[1, 2, 3, 4] Under these conditions, should a fault develop in the sheath, the escaping gas will protect the cable insulation from the effects of moisture until the fault can be located and rectified, even though the cable may be immersed in several feet of water.

Over the period for which such protective systems have been in use, a number of elegant methods have been devised for the final location of sheath faults, such as the use of radioactive tracer gases with Geiger counters[5] and halogenated organic compounds with halogen detectors,[6] but considerable savings in time and effort are possible by the intelligent use of elementary pneumatic methods based fundamentally on the establishment of a pressure gradient, at least for the initial rough location of the fault. In this way, the more precise methods of location may be rapidly brought to bear on the area in which the fault is known to lie, rather than on the whole of the faulty section of the cable, which might be several miles long. Nevertheless, the limitations of the pneumatic methods should be borne in mind, and a consideration of the factors influencing these limitations is the subject of this paper.

(2) THEORY

Giese[1] has considered that the flow of gas through a length of cable under steady-state conditions is analogous to the passage of electric current through a resistance, with pressure difference and rate of flow, respectively, taking the place of voltage and current in Ohm's law. This would result in the pressure decreasing linearly along the cable, the 'pneumatic resistance' of the cable being proportional to its length. Pech and Brune,[7] however, have shown experimentally that the pressure gradient under these conditions is not constant throughout the cable, and have derived an expression describing the flow of gas in the steady state, abandoning the pneumatic resistance per unit length as the fundamental parameter controlling flow in the cable in favour of a frictional coefficient.

An equivalent expression may be conveniently derived from the classical Poiseuille equation for the passage of gases through capillary tubes under steady conditions:

$$pv = \frac{(p_1^2 - p_2^2)\pi a^4}{16l\eta} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

If the interstitial paths in a non-coaxial type of telephone cable are considered to act as a number, n, of such capillary tubes of length kL, i.e. proportional to the cable length L, the formula may be written to express the passage of gas through the cable as

$$pv = \frac{n(p_1^2 - p_2^2)\pi a^4}{16kL\eta} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Therefore, in any given type of cable, where the number and size of the interstitial paths are constant and governed by the

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
Dr. Hooker is with Southern United Telephone Cables, Ltd.

number and size of the conductors, this expression may be reduced to

$$pv = \frac{p_1^2 - p_2^2}{RL} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where $$R = \frac{16k\eta}{n\pi a^4}$$

and is constant for a given cable passing any particular type of gas. This relationship explains the dependence of the viscosity of the gas used and the frictional coefficient which was observed by Pech and Brune in their work.

In a cable where the number of conductors is not too small, the number and effective diameter of the interstitial paths will be approximately proportional to the number and diameters of the conductors, respectively, and when a single working gas is being used, this last expression may be written as

$$R = \frac{K}{ND^4} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The constant K will change according to the make-up of the cable, i.e. star quad, unit twin, etc.

It is thought that the constant R might justifiably still be termed the pneumatic resistance per unit length, but its precise role in determining the rate of flow from the end pressures should be borne in mind.

(3) EXPERIMENTALLY DETERMINED VALUES OF R

In order to be able to test the theoretically derived dependencies of the constant R, its value was determined experimentally for a number of different cable types by measuring the rates of flow of air resulting from a range of pressures applied to the ends of various lengths of cable and using the relationship given in eqn. (3). Although it was not thought that the flow of air through coaxial cables would follow exactly the same law, owing to the larger paths being incapable of acting as capillaries, two types of cable containing coaxial pairs were included in the tests to enable an estimation of the effective values of R to be made under the same conditions of operation and to permit similar considerations of the limitations of the leak-location methods, to be treated in subsequent Sections, to be applied to them. In general, three sets of readings were taken for each cable sample, and the means of the results obtained from these tests are given in Table 1. In certain cases duplicate and triplicate tests on nominally identical cables, manufactured in different batches, were found to give values of R which were as much as 10% different from the mean. For convenience and to use units familiar in the field, the values of R were derived from pressures measured in pounds per square inch, lengths in yards and rates of flow in cubic centimetres per minute.

Table 1

MEAN VALUES OF R OBTAINED FOR A RANGE OF CABLE TYPES USING AIR AT A PRESSURE OF $37\frac{1}{2}$ LB/IN^2 ABSOLUTE

Cable	Type	Conductor diameter	R
		in	Units $\times 10^{-6}$
24 pair, 40 lb/mile* ..	Star quad	0·050	37·3
542 pair, 20 lb/mile ..	Star quad	0·035 5	13·4
400 pair, 20 lb/mile ..	Star quad	0·035 5	19·2
216 pair, 20 lb/mile ..	Star quad	0·035 5	20·2
28 pair, 10 lb/mile ..	Star quad	0·025	1 280
10 pair, 10 lb/mile ..	Star quad	0·025	1 615
76 pair, $6\frac{1}{2}$ lb/mile ..	Twin unit	0·020	100
{Two 0·375 in coaxial pairs +16 pair, 20 lb/mile	Coaxial +star quad	—	7·5
{Six 0·375 in coaxial pairs +4 pair, 25 lb/mile +344 pair, 20 lb/mile	Coaxial +star quad	—	2·3

* The gauges of copper conductors for telephone cables are referred to in terms of their nominal weight in pounds per mile.

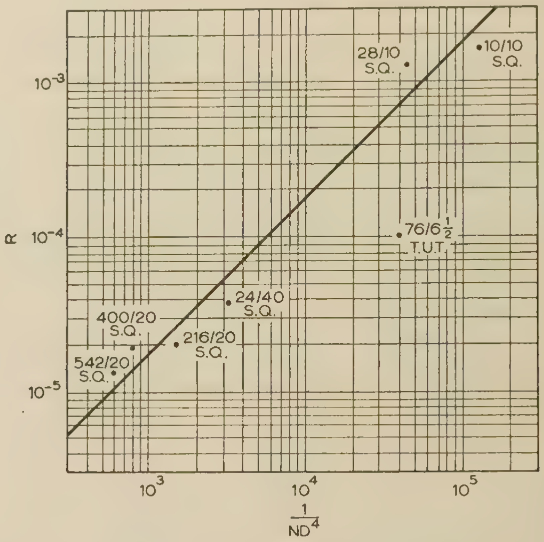


Fig. 1.—The relationship between the constant R and the factor $\frac{1}{ND^4}$ for a selection of cables.

S.Q. Star quad.
T.U.T. Twin-unit type.

Fig. 1 shows on a double logarithmic scale the values of R in Table 1 for the non-coaxial types of cable plotted against the reciprocal of the product (number of conductors) $\times$ (diameter)^4. It can be seen that the points for the star-quad cables fall closely around a line of unit gradient, showing that the dependency predicted theoretically does, in fact, exist. The position of the single value for the twin-unit-type specimen away from this line confirms that the value of K in eqn. (4) is different for star-quad and this latter type of cable.

(4) LIMITATIONS OF PNEUMATIC METHODS OF LEAK LOCATION

(4.1) Direct Pressure-Gradient Methods

When a pressurized cable develops a large enough leak, comparatively crude apparatus and techniques are capable of establishing, by methods described elsewhere, the existence of the pressure gradient in the cable with sufficient accuracy to permit determining the approximate position of the lowest pressure, and therefore the location of the fault itself. Nevertheless, it is of considerable value to be able to predict the accuracy of measurement or the sensitivity of the apparatus necessary to establish the gradients associated with smaller leaks, and it is in this direction that the numerical values of R and the relationships given in eqns. (3) and (4) are of use.

Although it has not been possible to derive a general expression for the non-steady condition of gas flow to enable the pressures within a cable at any time and at any position relative to the leak to be determined, certain approximations are possible which permit the conditions at low rates of flow to be estimated. Thus, if a small leak is postulated, such that the internal pressure falls but little during a given period, the rate of leak and therefore

gas flow within the cable will be substantially constant and eqn. (3) can be written as

$$pv = \text{constant} = \frac{p_1^2 - p_2^2}{RL} = \frac{(p_1 + p_2)(p_1 - p_2)}{RL} \quad . \quad (5)$$

This expression can be applied to an element of length of the cable dl where the mean pressure is p_m, so that

$$pv = \frac{2p_m}{R}\frac{dp}{dl} \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

from which the pressure gradient dp/dl along the cable can be calculated.

In practice, the approximate rate of leak of a system can be readily calculated from the rate of decrease in pressure at the ends and a knowledge of the air space of the cable. The rate of air flow at a point in the neighbourhood of the leak would be half the leak rate, since the cable on each side of the leak contributes towards its total value. If the value of R for the cable is known, the gradient to be expected can be determined.

If a manometer is used to measure the pressure at intervals along the length of the cable in order to establish the pressure gradient, the minimum reliable gradient which it will be possible to measure with simple apparatus will correspond to a pressure drop of about 2 mm head of the manometer fluid per measuring interval. Assuming the values of R given in Table 1 to hold at a working pressure of 10 lb/in^2 above atmospheric, in the case of the 542-pair, 20 lb/mile cable, a mercury manometer will just be able to detect the gradient produced by a leak of about 700 cm^3/min measured at atmospheric pressure for a measuring interval of 100 yd. If the same relationship is considered to apply to coaxial cables, a leak rate of over 1 litre/min will be the minimum producing a measurable gradient in the case of the six-coaxial-pair cable under similar conditions.

A considerable increase in sensitivity would be possible by using, say, water as the measuring fluid in the manometer, but under the working conditions described, this would necessitate a manometer of inordinate length. Nevertheless, by reducing the pressure in the cable to a value which could be read on a water manometer, it is possible to increase the sensitivity despite the fact that the leak rate also decreases. This can be demonstrated by assuming that the leak also obeys a similar law of gas flow, so that

$$pv_T = \frac{p_l^2 - p_A^2}{S} \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

and it can be shown that an increase in sensitivity by a factor of about 3 should result from decreasing the pressure from 10 to 1 lb/in^2 and using a water manometer instead of mercury.

It is also possible to increase the pressure gradient for a given leak and hence the ease of measurement, by injecting gas at a constant pressure at a single point. Under steady-state conditions, the leak will be fed from one side only, and on this side of the leak a gradient of twice the original will result, while on the other side the pressure will be constant along the length of the cable.

(4.2) Pressure-Sensitive Contactors

In a number of pressurized cable systems, pressure-sensitive contactors, installed at intervals along the length of the cable, are employed to indicate when the local pressure has fallen to a predetermined level due to leakage. The principle of the operation of such a system is closely connected with the setting up of a pressure gradient. It is intended that the contactor nearest a newly-formed leak will close first, and as the pressure-gradient front advances along the cable, other contactors at progressively increasing distances from the leak will close in that sequence. By collecting and timing signals from these contactors at an attended service point the approximate position of the leak may be estimated. Unfortunately, it is not possible to manufacture contactors to operate precisely identically, and two neighbouring contactors will, in fact, close at slightly different pressures. If the pressure gradient produced by a leak is smaller than that calculated from the difference in operating pressures of the contactors and their distance apart, it is possible for them to close in the wrong sequence, as shown in Fig. 2.

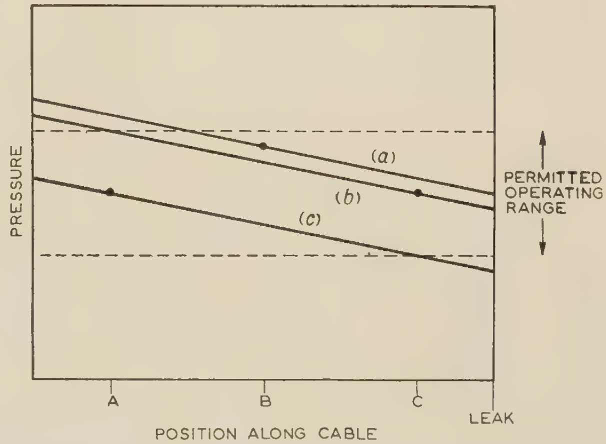


Fig. 2.—Explanation of possible maloperation of contactor system at one side of a small leak.

Desired order of operation of contactors C–B–A.
(a) First pressure condition operating B.
(b) Second pressure condition operating C.
(c) Third pressure condition operating A.
● Actual operating pressure.

The limiting flow rate which will permit the satisfactory use of a contactor system may be deduced from eqn. (6) as

$$pv = \frac{2p_m}{R}\frac{\Delta p}{l_c} \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

and it can be seen that such a failure of the system is more likely to occur on cables where R is lowest, as in coaxial types.

(4.3) Flow-Measuring Methods

By considering a length L of cable containing a leak at a distance l_A from one end, the ends of the cable being maintained at the same constant pressure p_c, it is possible to calculate from eqn. (3) the steady-state flow rate from each end towards the leak if this is at an unknown pressure p_l as follows:

Flow rate from end A.

$$pv_A = \frac{p_c^2 - p_l^2}{Rl_A} \quad . \quad . \quad . \quad . \quad . \quad (9a)$$

Flow rate from end B.

$$pv_B = \frac{p_c^2 - p_l^2}{R(L - l_A)} \quad . \quad . \quad . \quad . \quad . \quad (9b)$$

The simultaneous solution gives $\frac{pv_A}{pv_B} = \frac{L - l_A}{l_A}$ . . . (10)

Thus, by maintaining the ends of a leaking cable at the same pressure and measuring the rates of flow into the cable at each end, it is possible to calculate the position of the leak. This principle is well known,[8] and reports of its use for locating leaks in the sheaths of telephone cables date back over 20 years.[9] However, although methods are now available for measuring

flow rates without affecting the end pressures,[10, 11] the factor influencing the accuracy of location is now the closeness with which a given pressure may be set up independently at the ends of the cable, possibly several miles apart, and maintained at that value. It can be calculated that the error in location is given by the expression

$$\Delta l = \frac{2p_c(L - l_A)\delta p}{pv_T R(L - l_A) + 2p_c \delta p} \quad . \quad . \quad . \quad (11)$$

This relationship shows that errors in location depend on a number of factors, and, with other conditions being equal, the errors are worst for cables with low values of R. Nevertheless, it is possible to determine the accuracy of the pressure control necessary to locate a given leak to within an acceptable limit of error.

(4.4) Direction-Indicating Methods

It has been stated[4] that, in many cases, it is sufficient for a sheath fault to be located to a single jointed length of cable, since it may be preferable to replace a faulty length rather than repair it *in situ* under unfavourable conditions. As the faulty length will contain the point at the lowest pressure in the system, it will be unique in that gas will be flowing into it from both ends, and if the direction of flow could be established, a location would be possible on this basis.

The available methods of flow-direction indication require tappings into the cable at two points, preferably as far apart as possible, although, since such locations are often carried out in manholes, the points are generally a few yards distant at the most. One type of direction apparatus employs a sensitive differential manometer to indicate which tapping point of a pair is at the lower pressure, this being the one towards the leak. The limitations of the method are precisely those described earlier in Section 4.1 for the establishment of a pressure gradient. Although the precision and sensitivity of the pressure measurement can be much improved, the length of cable over which the measurements are made is considerably reduced and the relationship given in eqn. (6) still applies when the pressure gradient is calculated from these two factors.

An alternative form of direction indicator employs a by-pass path between the two tapping points, often with a transparent section of tubing into which a visible vapour or smoke is introduced as a marker, the observed direction of movement of this marker indicating the direction of the leak. For this apparatus to be effective, the resistance to flow offered by the complete by-pass path must be lower than, or at least comparable with, the resistance offered by the cable itself between the tapping points, so that sufficient gas travels via the parallel path to produce the required movement of the marker. It is obvious that such an arrangement is least effective with cables whose value of R is low, and this factor will govern the minimum flow rate for which the method may be used. However, an integrating by-pass direction indicator has been described[12] which employs the boundary between sections of dry and moist 'indicator' silica gel as a marker, and this device may be connected to a cable and left until a definite indication has been obtained, even with low rates of flow.

(5) CONCLUSIONS

It has been demonstrated that the flow of gas through pressurized telephone cables is largely governed by a factor depending on the number, size and arrangement of the conductors. This factor influences the pressure gradient produced in the cables by any particular sheath fault and affects the accuracy and sensitivity of the methods of leak location which are available. It also limits the smallest leak which may be successfully located by any of these methods. Since a given size of sheath fault will produce different pressure gradients in different cables, it is not possible to specify one method of location applicable to any particular magnitude of leak. Often, a combination of methods will be necessary before the final location is effected, the first indication being given by pressure-sensitive contactors or flow-measuring devices where either of these systems is installed, this being followed by a direct establishment of the pressure gradient using manometers at decreasing intervals along the cable until, when the change in pressure at successive points is too small to be measured reliably, the direction-indicating methods may be employed. The final location may use tracer gases or rely on visual examination of the suspect portion of the cable, aided by the old-established 'soap-suds' technique. It should be emphasized that all the methods examined above have their particular range of application and that consideration of the points indicated will help to predict the most useful method in any particular circumstance.

(6) ACKNOWLEDGMENTS

The author is indebted to the Director and General Manager of Southern United Telephone Cables, Ltd., for permission to publish the paper, and would like to thank Mr. L. H. Smith for his introduction to the problem of leak location in pressurized telephone cables.

(7) REFERENCES

(1) GIESE, R. G.: 'Gas Pressure for Telephone Cables', *Transactions of the American I.E.E.*, 1947, **66**, p. 471.

(2) ABERG, G.: 'Gas Control on Telephone Cables', *Ericsson Review*, 1952, **29**, p. 108.

(3) CAZAUX, A.: 'La maintenance des câbles par pression gazeuse. 2e Partie: La maintenance par pression des câbles téléphoniques à grand distance', *Câbles et Transmission*, 1950, **4**, p. 343.

(4) KEEP, J. F.: 'Field Trial of Pressurised Cable System', *Post Office Electrical Engineers' Journal*, 1955, **48**, p. 7.

(5) GUÉRON, J., and PAGÈS, A.: 'Expériences de localisation de fuites dans des câbles téléphoniques souterains au moyen de gaz radioactif traceur', *Câbles et Transmission*, 1952, **6**, p. 96.

(6) BOWKER, M. W.: 'Finding Gas Leaks in Cable Sheaths', *Bell Laboratories Record*, 1952, **30**, p. 359.

(7) PECH, H., and BRUNE, D.: 'La maintenance des câbles par pression gazeuse. 1re Partie: Essai théorique sur les phénomènes pneumatiques dans un câble téléphonique parcouru par un courant gazeux', *Câbles et Transmission*, 1950, **4**, p. 336.

(8) HOCHSTADTER, M., VOGEL, W., and BOWDEN, E.: 'The Pressure Cable', *Journal of the Royal Society of Arts*, 1931, **80**, p. 84.

(9) BUSS, K., and VOGEL, W.: 'Eine nichtelektrische Methode zur Bestimmung von Bleimantelfehlen in Telephonkabeln', *Europäischer Fernsprechdienst*, 1935, **41**, p. 235.

(10) HENCKEL, A.: 'Gas Control on Telephone Cables by Gas Flow Meters and Automatic Pressure Compensation', *Ericsson Review*, 1954, **31**, p. 34.

(11) British Patent No. 734298.

(12) CHAVIGNER, A.: 'La maintenance des câbles par pression gazeuse. 4e Partie: La localisation sur le terrain des fuites d'un câble maintenu sous pression', *Câbles et Transmission*, 1951, **5**, p. 32.

:1.317.34: 621.372.2

The Institution of Electrical Engineers
Paper No. 2689 R
Sept. 1958
©

MEASUREMENT OF IMPEDANCE AND ATTENUATION OF A CABLE THROUGH AN ARBITRARY LOSS-FREE JUNCTION

By J. ALLISON, B.Sc.(Eng.), Ph.D., Graduate, and F. A. BENSON, D.Eng., Ph.D., Associate Member.

(*The paper was first received* 26*th July*, 1957, *in revised form* 16*th January, and in final form* 10*th May*, 1958.)

SUMMARY

The paper considers the problem of finding the impedance and attenuation of a transmission line when measured through an arbitrary loss-free junction. Several possibilities for the exact determination of impedance in such a case are mentioned, and the results of tests carried out to determine the usefulness and accuracy of certain methods are presented. Some information is also given on the experimental accuracy of the well-known circle-diagram technique for determining transmission-line characteristics.

LIST OF PRINCIPAL SYMBOLS

α = Attenuation coefficient.
β = Phase-change coefficient.
ε = Reflection coefficient of cable and terminating reactance.
ρ = Input reflection coefficient.
l = Length of transmission line.
λ = Wavelength along the line.
Z_0 = Image impedance of transmission line.

(1) INTRODUCTION

The work described here arose from a study of the effects of irregularities in coaxial cables and measurement of the attenuations produced by such cables at microwave frequencies. Some information about irregularities in cables may be obtained by measuring the variations of characteristic impedance with frequency, as pointed out by Blackband.[1] The geometric mean of open- and short-circuit impedances, as given by the circle diagram and two-point methods described below, provides a value for the image impedance from one end of the cable. The cable will behave as an asymmetrical four-pole network, because of the random position of its irregularities, and will thus have differing image impedances. The characteristic impedance of the cable is then the geometric mean of the image impedances measured from each end.

At the frequencies under consideration, measurements of both impedance and attenuation are commonly made using an air-spaced coaxial standing-wave indicator. If a cable is terminated by a length of air-spaced line containing a sliding short-circuit, the input-admittance values ($Y = G + jB$) when plotted on an Argand diagram, trace out a circle as the short-circuit is moved over half a wavelength. The attenuation and image impedance of the cable can be found from this circle. The procedure is, however, very laborious when values at many wavelengths are required, and for this reason a more rapid technique has been devised by Blackband and Brown[2] which is generally referred to as the two-point method. The impedances of the cable sample can also be less conveniently measured by a method described by Cook.[3]

As pointed out by Oliver,[4] in order to obtain accurate results in measurements of this kind without applying corrections, it is necessary, not only for the standing-wave indicator itself to be uniform, but also for the uniformity to extend along the whole length of line between the indicator and the unknown impedance. This condition is not easily attainable in practice because, unfortunately, some form of adapter section is required to couple the indicator to the cable which usually introduces a discontinuity.

Any loss-free discontinuity between the measuring line and the cable will cause serious errors in the measured value of impedance, but the circle-diagram method still provides an accurate value for line attenuation. In this paper various possibilities for the exact determination of impedance when measured through a discontinuity are mentioned, and the results of experiments carried out to determine the usefulness and accuracy of some of the methods are presented. Information is also given on the experimental accuracy of the circle-diagram technique for determining cable characteristics, including the two-point method due to Blackband and Brown.[2]

(2) DISCONTINUITIES AT THE JUNCTION BETWEEN MEASURING LINE AND CABLE SAMPLE

Discontinuities can occur in the measuring system due to coaxial-line eccentricity, reflections from the slot, dielectric supports for the inner conductor, mismatch at the extremities of the conical adaptor, steps at the inner conductor junctions and geometric irregularity at braiding junctions.[4] Although it is possible to reduce the effects of some of these discontinuities by suitable design, it is very difficult to eliminate all mismatch, and so it is necessary to have some method of deducing the effect that a particular junction will have no measurements made through it.

A convenient method for describing the behaviour of the junction is by means of an equivalent circuit. Whinnery *et al.*[5] have shown that the effect of step-type discontinuities on coaxial systems can be calculated from an equivalent circuit in which local waves existing at the change of section are accounted for by a lumped admittance shunted between the lines at the junction. Such calculations are applicable only in very special cases, as normally the discontinuities are not of this form. More general lossless discontinuities may be represented by a lossless four-terminal network requiring three parameters for its complete specification. There is, of course, no unique equivalent circuit for a particular junction, but an infinite number of them, and although a network of lumped circuit elements may be chosen which exactly represents a junction at one specific frequency, such equivalent circuits are purely artificial devices, and the parameters vary with frequency in an arbitrary way.

The properties of the four-terminal network are often represented by the input and output currents and voltages as in Fig. 1(*a*), by an equivalent T-section as at (*b*) or by an equivalent π-section as at (*c*).[6, 7] In each case the sign and direction convention for voltages and currents are different. The relationships between the input and output voltages and currents in Fig. 1, involving the *A*, *B*, *C* and *D* parameters and often written in the matrix form, are well known. A scattering matrix (Fig. 2)

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

Dr. Allison and Dr. Benson are in the Department of Electrical Engineering, University of Sheffield.

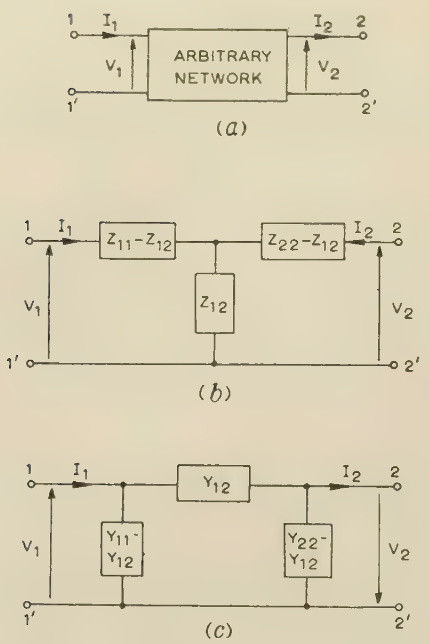


Fig. 1.—Representation of junction by impedance and admittance networks.

Fig. 2.—Scattering-matrix representation of junction.

may also be used to specify a junction, the incident and reflected waves of the four-pole network then being given by

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \quad . \quad . \quad . \quad . \quad (1)$$

where
$$S \equiv \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} \quad . \quad . \quad . \quad . \quad . \quad (2)$$

is the scattering matrix.

Relationships between elements of impedance, admittance and scattering matrices and A, B, C and D coefficients can easily be found, and, as mentioned already, because of reciprocity, only three parameters are required to specify the properties of the network.

A further way of defining a lossless discontinuity is by two reflection coefficients and one transmission coefficient at any specific frequency.[20,22]

Several methods are now available for determining the equivalent network or scattering-matrix elements of a four-terminal structure. The best-known method is probably that described by Deschamps,[8–11] which is based on the measurement of the reflection coefficient $\rho = |\rho|\varepsilon^{j\phi}$ at terminals 1, 1′ when terminals 2, 2′ are connected to a line of variable length l terminated in a short-circuit. Then

$$\rho = [(s_{12}^2 - s_{11}s_{22})\rho_L + s_{11}]/[1 - s_{22}\rho_L] \quad . \quad . \quad (3)$$

where
$$\rho_L = a_2/b_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Eqn. (3) is of the well-known form

$$w = (az + b)/(cz + d) \quad . \quad . \quad . \quad . \quad (5)$$

For such a transformation, circles in the z-plane (ρ_L-plane appear as circles in the w-plane (ρ-plane). Deschamps has derive graphical procedures for evaluating the magnitudes and angle of s_{11}, s_{12} and s_{22} by using non-Euclidian-geometry theorem and the properties of bilinear transformations. Simplifie proofs based on plane geometry have been derived by Store Sheingold and Stein,[12] who also show how impedance may b measured through a junction. Further information on graphica analyses can be found in papers by Dukes,[13, 14] Altschuler an Felsen,[15] and elsewhere.[7, 16] Some of the more common method of defining the equivalent circuit of a junction which are c importance for microwave applications, including one arrangement which results from the well-known Weissfloch transforme theorem,[18, 19] have been given by Montgomery *et al.*[6]

(3) MEASUREMENT OF CABLE ATTENUATION THROUGH A DISCONTINUITY

By representing the discontinuity by two reflection coefficient and one transmission coefficient it is shown in Section 9.1 tha the input reflection coefficient ρ of the load when measure through a junction is a bilinear transformation of the reflectio coefficient ρ_L of the cable and terminating reactance. Thus, if series of measurements of ρ is taken, as the terminating reactanc is varied from $+\infty$ to $-\infty$, then, since ρ_L describes a circle, s does ρ. The attenuation of the cable is given by the followin expression (see Appendix 9.1):

$$\alpha = \frac{1}{2l} \operatorname{arc\,cosh} \{[1 + (r')^2 - |s|^2]/2r'\} \quad . \quad . \quad (6$$

where r' is the radius of the ρ circle and $|s|$ is the distance of it centre from the origin of the Smith chart. The value of α, foun by measuring r' and $|s|$ from the ρ-circle and using eqn. (6), i true, irrespective of any discontinuity which may exist betwee the measuring line and cable sample.[4] The well-known formul for calculating attenuation, which is derived by assuming ther are not discontinuities, is in fact

$$\alpha = \frac{1}{l} \operatorname{arc\,tanh} [(r_{min}/r_{max})^{1/2}] \quad . \quad . \quad . \quad . \quad (7$$

where r_{max} and r_{min} are the maximum and minimum values respectively, for the resistive component of the impedance o the cable plus variable reactance, when measured through th junction. It is shown in Appendix 9.2 by a novel method tha eqns. (6) and (7) are equivalent.

(4) IMPEDANCE MEASUREMENTS THROUGH A DISCONTINUITY

The image impedance of the cable sample, Z_0, is commonl found from the measurements discussed in Section 3 using th formula

$$Z_0 = Z_0'(r_{min}r_{max})^{1/2} \quad . \quad . \quad . \quad . \quad . \quad (8$$

where Z_0' is the characteristic impedance of the measuring line This gives the wrong value for Z_0 if no allowance is made fo discontinuities,[4] as is also clear from Appendix 9.2. Method of taking account of discontinuities in such measurements hav been discussed, for example by Oliver.[4]

To investigate the possibility of the exact determination of Z when measured through a discontinuity, using the same experimental data as in Section 3, consider eqn. (17) [Appendix (9.1)] This equation may be rewritten as

$$|\rho|\varepsilon^{j\phi} = |\rho_1|\varepsilon^{j\phi_1} + [(1 - |\rho_2|^2)\varepsilon^{j(\phi_1+\phi_2\pm\pi)}|\rho_L|\varepsilon^{j\psi}]/[1 - |\rho_2|\varepsilon^{j\phi_2}|\rho_L|\varepsilon^{j\psi}] \quad (9$$

The values of $|\rho_2|$ and phase angles ϕ_1 and ϕ_2, which completely define the discontinuity, can be obtained either by Feenberg's method[21] or from measurements of the input impedance when the junction is open-circuited and short-circuited at a reference plane and also $\lambda/8$ from this plane. Thus it is possible to substitute these values in eqn. (9) and so transform all the measured values of input impedance to give the exact reflection-coefficient circle for the cable plus terminating reactance. This process, when repeated for every point on the original input-impedance circle, is, however, very tedious, and simpler methods were investigated.

(4.1) T-Section Representation of the Junction

When the equivalent network parameters of the junction are known, the impedance transformation through the network may be found by using conventional circuit calculations, although these may be tedious. In nearly all cases, however, one is not interested in determining the equivalent circuit of the discontinuity. Direct transformation relationships between the input and load impedances prove to be much more useful if these relationships are available in a simple form, as pointed out by Mittra.[17]

Methods[4, 25] exist for calibrating a slotted line for measuring impedance through a discontinuity. M. H. Oliver's method[4] is suitable for lossless structures only, and the technique described by A. A. Oliner[25] involves approximations for lossy junctions. Both the methods require a large number of measurements. The method described here, where the junction is represented by a T-section similar to Fig. 1(*b*), does not involve any approximation even for large discontinuities or for those with appreciable loss. Only a small number of measurements are required in the calibration procedure and the treatment of the results is simple. In fact, the calibration constants are obtained directly for a lossless discontinuity by plotting measured values of reflection coefficient on a Smith chart. When the structure is lossy, some simple graphical constructions have to be carried out.

It is shown in Appendix 9.3 that, when the circuit of Fig. 1(*b*) is terminated by an impedance z_l, there is a linear relationship between the reflection coefficient ρ at the input to the circuit and another subsidiary reflection coefficient ρ''. Therefore ρ'' describes a circle as the reactive component of z_l is varied. Now consider the two planes ρ and ρ''. If the load is a pure variable reactance, i.e. $z_l = jx_l$ and $r_l = 0$, then, as x_l varies, the outer circle of the Smith chart is obtained on the ρ''-plane. If ρ is measured under these conditions a circle is obtained on the ρ-plane (Fig. 3). When the output is on open-circuit, i.e. when

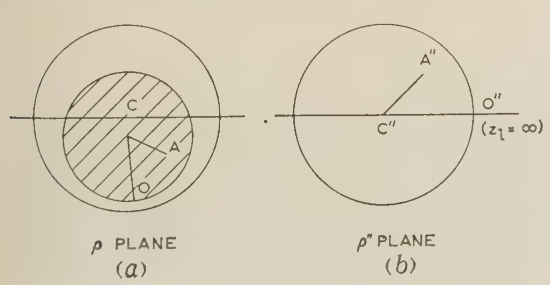


Fig. 3.—Input-reflection-coefficient plane and subsidiary reflection-coefficient plane.

$z_l = \infty$, point O'' on the ρ''-plane is obtained. The corresponding point on the ρ-plane will be on the circumference of the shaded circle (centre C) at O. Comparison of CO and C''O'' gives the transformation relation between ρ and ρ'' in both magnitude and phase. Thus, in general, transformation of a point A to a point A'' can be effected as follows:

$$CA/C''A'' = CO/C''O'' \quad \text{and} \quad \angle\underline{ACO} = \angle\underline{A''C''O''}$$

In the specific case of a lossless junction, the perimeter of the shaded circle on the ρ-plane is also the outer rim of the Smith chart, and the transformation of points from the ρ-plane to the ρ''-plane merely involves an angular displacement of the points, i.e. $CO/C''O'' = CA/C''A'' = 1$. Thus, if ρ_L is measured for a certain load impedance, point A on the ρ-plane is obtained. This transforms to the point A'' on the ρ''-plane, which has components r''_A and x''_A. It is shown in Appendix 9.3 [eqn. (49)] that the load impedance z_{lA} is given by the expression

$$z_{lA} = r'r''_A + jr'x''_A - jx'$$

where r' and x' are constants which can be found from measurements on the short-circuited junction. Hence, all points on the measured impedance circle can be transformed to give the impedance circle for the load.

(4.2) Representation of the Junction by Two Lengths of Transmission Line and a Shunt Capacitance

It is sometimes more convenient to represent the transition region between the measuring line and the cable by the four-terminal network of Fig. 4. The junction is completely defined

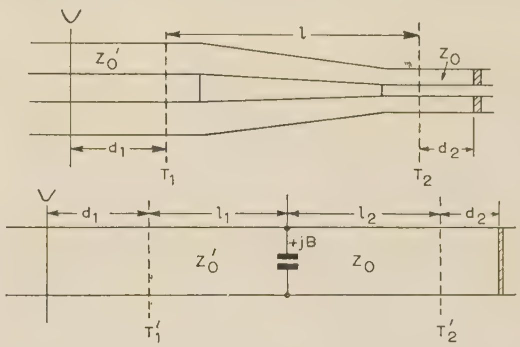


Fig. 4.—Representation of junction by two lengths of transmission line and a shunt susceptance.

by the two lengths of line l_1 and l_2, having characteristic impedances equal, respectively, to those of the two lines being joined, and a shunt capacitance. When the parameters are expressed in this form it is relatively simple, from input-impedance measurements through the junction, to calculate the value of the unknown impedance using Smith-chart methods.

The parameters may be determined from data obtained by Feenberg's method.[21] An advantage of this is that errors in locating nodal positions relative to the fixed reference planes tend to be reduced. The method consists in measuring the distance d_1 of a voltage node in the measuring line from an arbitrary reference plane T_1 when a short-circuiting piston, distant d_2 from a second arbitrary reference plane T_2 at the opposite side of the junction, is moved through at least half a wavelength.

When $(d_1 + d_2)$ is a maximum $d_1 + l_1 = n\lambda/2$, and $d_2 + l_2 = m\lambda/2$, where m and n are integers. Thus l_1 and l_2 can be calculated. Also, when $(d_1 + d_2)$ is a minimum, B can be determined as it is equal to the conjugate of the total susceptance of the two short-circuited lines of lengths $(l_1 + d_1')$ and $(l_2 + d_2')$ in parallel. It may be shown that

$$(jB)^* = -2j\cot\left[2\pi(l_1 + d_1')/\lambda\right] \quad . \quad . \quad . \quad (10)$$

The junction parameters having been obtained, the load impedance can be determined in the following way. First, the input admittance Y_i is measured when the junction and cable are terminated by a variable reactance. It is shown in Appendix 9.4 that, if the input to the equivalent circuit of the junction is closed by an admittance Y_i^*, the admittance looking in from the load end of the network is Y_r^*, the conjugate of the total load admittance. The measured input admittances can thus be transformed to give the admittance circle of the load alone by normal Smith-chart techniques, as indicated in Section 9.4.

(5) EXPERIMENTAL STUDIES

(5.1) Admittance Circles for Measurement of Attenuation and Impedance

The common method of constructing admittance circles, which has been employed during the present investigations, consists of locating a minimum with respect to a reference plane and determining the voltage standing-wave ratio at this point to provide one value of admittance on a Smith chart.

Movement of the short-circuiting piston through half a wavelength causes the measured values to trace out an admittance circle on the chart.

Measurements were made on a sample of UR66 cable to determine the experimental accuracy of this technique. The attenuation of the cable was determined with varying piston ranges giving the results shown in Table. 1. It is evident from

Table 1

MEASURED ATTENUATION FOR VARYING PLANE OF TERMINATING REACTANCE

Initial piston position	Attenuation
cm	dB/ft
3·0	0·405
3·8	0·403
4·7	0·391
5·4	0·395
	Mean value: 0·399

these figures that the repetition accuracy of admittance-circle measurements is of the order of $\pm 2\%$.

The circles drawn for these preliminary measurements have been examined to determine the probable errors which would have arisen had Blackband and Brown's two-point method[2] been employed to determine the attenuation. Values obtained for the attenuation, found by assuming that the zero susceptance points can be accurately located, are shown in Table 2. The

Table 2

ATTENUATION AS FOUND BY TWO-POINT METHOD

Initial piston position	Attenuation
cm	dB/ft
3·0	0·426
3·8	0·423
4·7	0·413
5·4	0·426
	Mean value: 0·422

results indicate a difference of 5% between the value obtained for the attenuation by the two-point method and that obtained using a complete admittance circle. Of course, in practice, the zero-susceptance points may not be so accurately located. This could lead to differences greater than 5% between the two methods.

(5.2) Measurements of Attenuation and Impedance through a Discontinuity

Some impedance measurements have been made through discontinuities at 10 Gc/s. The normal junction (Fig. 4) as used for cable measurements was replaced by a junction which had an additional discontinuity consisting of a 2 B.A. screw which

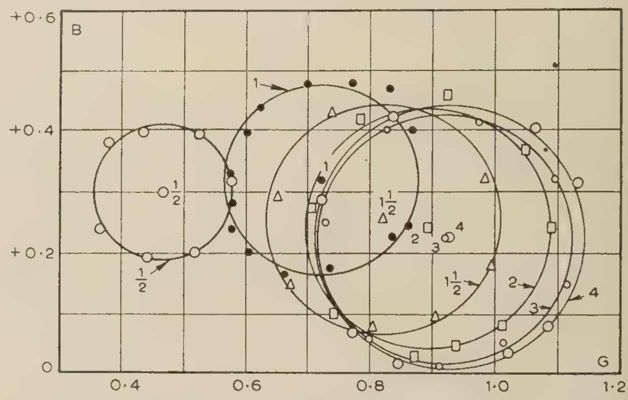


Fig. 5.—Admittance circle for cable UR32 measured through junction with screw insert.

could be screwed radially into the junction. Admittance circles were then determined for various positions of the screw. The resulting circles are shown in Fig. 5 and lead to the apparent values of impedance shown in Table 3. Attenuations have also been calculated and are included in that Table to confirm that

Table 3

ATTENUATION AND IMAGE IMPEDANCE AS MEASURED THROUGH A DISCONTINUOUS JUNCTION

Circle No.	Turns of screw out from inner conductor	Attenuation	Line impedance*
		dB/ft	
1	½	1·18	2·22
2	1	1·25	1·43
3	1½	1·24	1·25
4	2	1·24	1·15
5	3	1·24	1·12
6	4	1·21	1·12

* Calculated from eqn. (8) normalized with respect to Z_0'.

the technique provides a constant attenuation within the expected experimental accuracy, in spite of the presence of a discontinuity. The image impedances vary very considerably and are entirely dependent on the discontinuity present at the junction.

(5.3) Measurement of Impedance through a Junction

(5.3.1) T-Network Representation Theory.

A transmission line with a known impedance has been examined using the method described in Section 4.1 to ascertain the usefulness of this technique and the accuracy with which impedance can be measured through a junction.

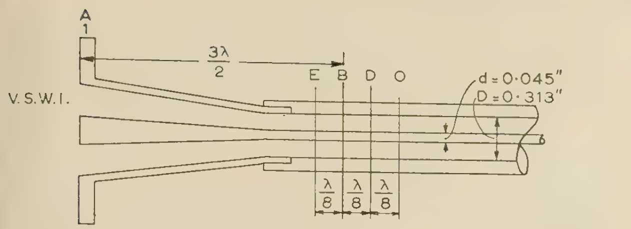


Fig. 6.—Junction between standing-wave indicator and air-filled line.

An air-filled coaxial line was connected to the measuring line via a tapered conical junction, as illustrated in Fig. 6. The line had dimensions as shown, and the resistivities of the outer and inner conductors were $6 \cdot 5 \times 10^{-6}$ and $1 \cdot 78 \times 10^{-6}$ ohm-cm, respectively. These values correspond to a characteristic impedance of approximately 110 ohms and an attenuation of $0 \cdot 1$ dB/ft at a frequency of 10 Gc/s.

An impedance circle was obtained for a 3 ft length of the line when terminated by a variable short-circuiting piston, i.e. circle A in Fig. 7.

The line length was then very much shortened, so that the total loss of the line plus junction was negligible, and the minimum was located when a short-circuit was placed at points

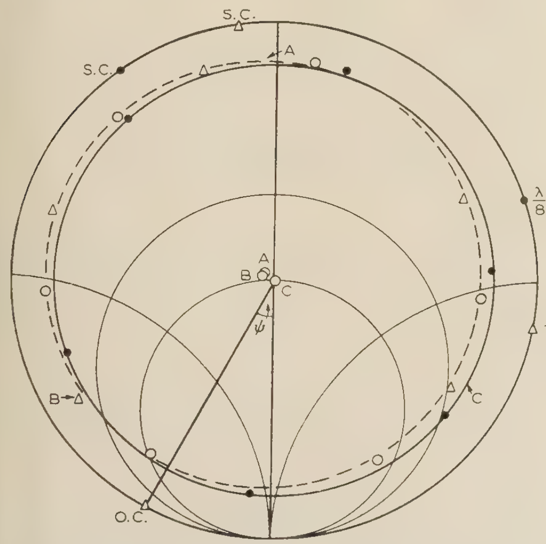


Fig. 7.—Determination of impedance when measured through a junction by the method described in Section 5.3.1.

A—Input impedance circle.
B—Impedance circle of cable sample.

B, O, D and E (Fig. 6). The distance of B from the reference plane at A was made an integral number of half-wavelengths. Point O, then, is equivalent to an open-circuit and D to a length of short-circuited line with respect to the reference plane at B. These points are therefore identical with those of the same letter on the diagram for the ρ-plane (Fig. 12).

The procedure for calculating the characteristic impedance of the line from these data is then as described in Section 4.1 and Appendix 9.3. All points on the original circle (A) are then transformed to the ρ''-plane and the points (B) on the resulting circle (Fig. 7) provide values for r''_A and x''_A. These values, together with those for r' and x', enable the complex load impedance to be calculated for each point on the original experimental circle using eqn. (49). The final circle as described by these points, C in Fig. 7, is the impedance circle for the line and is independent of the discontinuity at the junction.

It will be noted that the load-impedance circle has its centre at the origin of the Smith chart when correction is made for junction effects. Thus the intercepts of this circle on the resistive axis (r_{max} and r_{min}) are reciprocal. Also, since the centre now lies on the R-axis, the usual formula

$$Z_0 = Z'_0(r_{max}r_{min})^{1/2}$$

is exact; hence $Z_0 = Z'_0$. In other words, the impedance of the air-filled line as found by this method is equal to unity when referred to the very short length of the same line, as would be expected. However, the difficulty of choosing the best circle through the transformed points on the Smith chart, together with experimental errors, indicate that a value could be obtained for the line impedance which was 6% inaccurate. The original circle, on the other hand, provides a value for the line impedance which is over 18% inaccurate for this particular junction.

(5.3.2) Two Lengths of Transmission Line and Shunt Susceptance Representation Theory.

The usefulness and accuracy of the method described in Section 4.2 for measuring impedances through a junction have been examined, again using an air-filled line.

First, the impedance circle of the same 3 ft length of air-filled line when measured through the discontinuity was determined

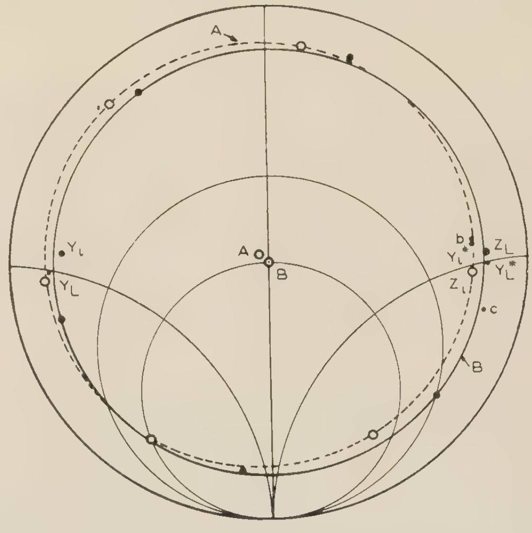


Fig. 8.—Determination of impedance when measured through a junction by the method described in Section 5.3.2.

A—Input impedance circle (ρ-plane).
B—Points on circle transformed to ρ″-plane.
C—Impedance circle of cable sample.

as before (Z_i-circle of Fig. 8). Then, instead of locating a node for specific short-circuit positions as was done in the previous experiment, nodes were located as the short-circuit was moved through $\lambda/2$ in the line, the length of which was again such as to make losses negligibly small.

The results of this experiment are plotted in Fig. 9. From this graph and using eqn. (10), the parameters of the junction are found to be $l_1 = 0 \cdot 498\lambda$, $l_2 = 0 \cdot 483\lambda$, $B = 0 \cdot 354$. It is then possible, using the Smith-chart technique as described in Appendix 9.4, to transform the original Z_i-circle to the true load-impedance circle for the air-filled line (Z_L-circle of Fig. 8).

The resulting circle is centred at the origin of the chart, indicating that the characteristic impedance of the measured line is identical with that of line l_2. As before, the overall accuracy of the measurement is of the order of 6%.

Again, if the effect of discontinuity at the junction had been ignored, errors as great as 18% could be encountered in the

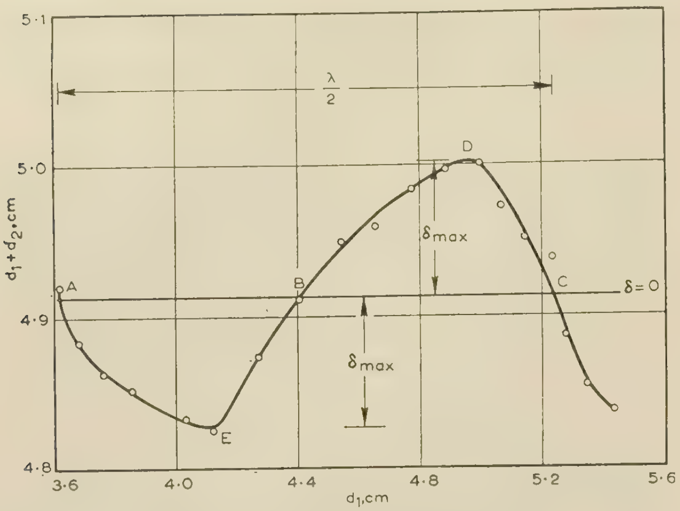


Fig. 9.—Experimentally observed curve of d_1 versus $d_1 + d_2$ for the conical junction (Feenberg's method).

evaluation of Z_0. On the other hand, both circles (Z_i and Z_L) yield the same value for the attenuation of the line, as was anticipated.

(6) CONCLUSIONS

A number of possibilities for the exact determination of impedance when measured through a discontinuity have been discussed, and the usefulness and accuracy of some of the methods presented have been studied experimentally. Unfortunately, such exact measurements would be very laborious when readings at many wavelengths are required, and it is highly desirable in such cases to concentrate on eliminating discontinuities so that the simple two-point method can be adopted and relied upon to give satisfactory results.

(7) ACKNOWLEDGMENTS

The work recorded here was carried out at the University of Sheffield, and thanks are due for the facilities afforded in the laboratories there. The authors are indebted to Professor A. L. Cullen for very helpful comments on the theory, and to Mr. B. C. Williams for his very kind assistance with some of the experimental work. They also wish to thank Mr. W. T. Blackband for his many helpful suggestions and very stimulating discussions.

The investigation was carried out under a contract placed by the Ministry of Supply, and thanks are given for their permission to publish the paper.

(8) REFERENCES

(1) Blackband, W. T.: 'A Graphical Approach to the Study of Irregularities in Transmission Lines', Royal Aircraft Establishment Internal Report, February, 1956.

(2) Blackband, W. T., and Brown, D. R.: 'The Two-Point Method of Measuring Characteristic Impedance and Attenuation of Cables at 3 000 Mc/s', *Journal I.E.E.*, 1946, **93**, Part IIIA, p. 1383.

(3) Cook, A. H.: 'The Measurement of Attenuation at X-Band Frequencies by Impedance Circle Methods', Admiralty Signal and Radar Establishment Report, XRC5/45/1, 1945.

(4) Oliver, M. H.: 'Discontinuities in Concentric-Line Impedance-Measuring Apparatus', *Proceedings I.E.E.*, Paper No. 884 R, January, 1950 (**97**, Part III, p. 29).

(5) Whinnery, J. R., Jamieson, H. W., and Robbins, T. E.: 'Coaxial Line Discontinuities', *Proceedings of th Institute of Radio Engineers*, 1944, **32**, p. 695.

(6) Montgomery, C. G., Dicke, R. H., and Purcell, E. M. (Eds.): 'Principles of Microwave Circuits' (McGraw-Hill, 1948).

(7) King, R. W. P.: 'Transmission Line Theory' (McGraw-Hill, 1955).

(8) Deschamps, G. A.: 'Geometric Viewpoints in the Repre sentation of Waveguides and Waveguide Junctions', Proceedings of Symposium on Modern Network Syn thesis, New York, September, 1952, p. 277.

(9) Deschamps, G. A.: 'Determination of Reflection Coeffi cients and Insertion Loss of Waveguide Junctions', *Journal of Applied Physics*, 1953, **24**, p. 1046.

(10) Deschamps, G. A.: 'New Chart for the Solution of Trans mission-Line and Polarization Problems', *Electrica Communication*, 1953, **30**, p. 247.
See also *Transactions of the Institute of Radio Engineers Professional Group on Microwave Theory and Techniques*, 1953, **1**, p. 5.

(11) Deschamps, G. A.: 'Geometrical Representation of th Polarization of a Plane Electromagnetic Wave', *Pro ceedings of the Institute of Radio Engineers*, 1951, **39**, p. 540.

(12) Storer, J. E., Sheingold, L. S., and Stein, S.: 'A Simpl Graphical Analysis of a Waveguide Junction', *ibid.*, 195 **41**, p. 1004.

(13) Dukes, J. M. C.: 'Waveguides and Waveguide Junctions *Wireless Engineer*, 1955, **32**, p. 65.

(14) Dukes, J. M. C.: 'Transmission-Line Termination', *ibid.*, 1955, **32**, p. 266.

(15) Altschuler, H. M., and Felsen, L. B.: 'Network Method in Microwave Measurements', Proceedings of the Sym posium on Modern Advances in Microwave Tech niques, Polytechnic Institute of Brooklyn, November 1954.

(16) Storer, G. E., Sheingold, L. S., Stein, S., Felsen, L. B. and Oliner, A. A.: Comment on 'A Simple Graphical Analysis of a Two-Port Waveguide Junction', *Pro ceedings of the Institute of Radio Engineers*, 1954, **42**, p. 1447.

(17) Mittra, R.: 'Impedance Measurements through a Dis continuity in a Transmission Line', *Institute of Radi Engineers Convention Record*, 1955, Part 8, p. 85.

(18) Weissfloch, A.: 'Ein Transformationssatz über Verlustlos Vierpole und seine Anwendung auf die experimentell Untersuchung von Dezimeter- und Zentimeter-Wellen Schaltungen', *Hochfrequenztechnik*, 1942, **60**, p. 67.

(19) Weissfloch, A.: 'Schaltungstheorie und Messtechnik de Dezimeter- und Zentimeter-Wellengebietes' (Verla Birkhauser, 1954).

(20) Cullen, A. L.: 'Measurement of Microwave Transmissio Efficiency', *Wireless Engineer*, 1949, **26**, p. 255.

(21) Feenberg, E.: 'Relation between Nodal Positions an Standing-Wave Ratio in a Complex Transmissio System', *Journal of Applied Physics*, 1946, **17**, p. 530.

(22) Barlow, H. M., and Cullen, A. L.: 'Microwave Measure ments' (Constable, 1950).

(23) Schelkunoff, S. A.: 'Electromagnetic Waves' (Va Nostrand, 1943).

(24) Jackson, W.: 'High Frequency Transmission Lines' (Methuen, 1945).

(25) Oliver, A. A.: 'The Calibration of the Slotted Section fo Precision Microwave Measurements', *Review of Scientifi Instruments*, 1954, **25**, p. 13.

(9) APPENDICES

(9.1) Calculation of Attenuation from Smith-Chart Data

The reflection coefficient referred to Z_{02} (Fig. 10) is

$$\rho_L = \varepsilon^{-2\alpha l}\varepsilon^{-j2\beta l}\varepsilon^{-j\phi} \quad . \quad . \quad . \quad . \quad (11)$$

where $\varepsilon^{-j\phi}$ is the reflection coefficient of the terminating reactance X, referred to Z_{02}. If the terminating reactance is varied, ρ_L describes a circle of radius

$$r = \varepsilon^{-2\alpha l} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

in the complex reflection-coefficient plane.

If the junction behaves as an ideal transformer of turns ratio $(Z_{01}/Z_{02})^{1/2}$, the reflection coefficient in the terminal plane ρ referred to Z_{01} is

$$\rho = \rho_L \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

If the junction is an arbitrary loss-free transition, it can be completely specified by two reflection coefficients and one transmission coefficient (Fig. 10).

Then[22]
$$1 - |\rho_1|^2 = |\tau|^2 = 1 - |\rho_2|^2 \quad . \quad . \quad . \quad (14)$$

Also
$$\theta = \frac{\phi_1 + \phi_2}{2} \pm \frac{\pi}{2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

where
$$\rho_1 = |\rho_1|\varepsilon^{j\phi_1},\ \rho_2 = |\rho_2|\varepsilon^{j\phi_2} \text{ and } \tau = |\tau|\varepsilon^{j\theta} \quad (16)$$

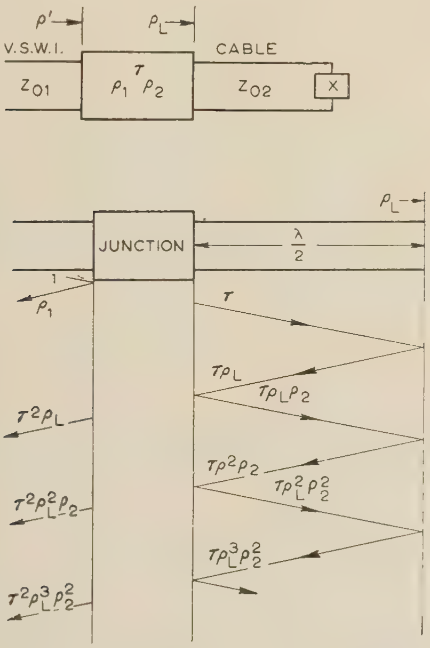


Fig. 10.—Total reflections from a junction.

Reflections caused by the junction are shown in Fig. 10. The resulting reflection coefficient referred to the ρ-plane is then

$$\rho = \rho_1 + \tau^2\rho_L[1 + \rho_L\rho_2 + (\rho_L\rho_2)^2 \ldots]$$

or
$$\rho = \rho_1 + \frac{\tau^2\rho_L}{1 - \rho_L\rho_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

Thus ρ is a bilinear transformation of ρ_L, so that if ρ_L describes a circle, ρ also describes a circle. To find the radius (r') of the circle described by ρ, we first find the radius of the circle described by w; hence

$$w = \frac{\tau^2\rho_L}{1 - \rho_L\rho_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

The two radii will be equal, since ρ_1 in eqn. (17) merely represents a shift in position. The equation satisfied by ρ_L may be written[23]

$$\rho_L\rho_L^* = \varepsilon^{-4\alpha l} = r^2 \quad . \quad . \quad . \quad . \quad (19)$$

From eqn. (18) we have
$$\rho_L = \frac{w}{\tau^2 + w\rho_2} \quad . \quad . \quad . \quad . \quad (20)$$

Hence eqns. (19) and (20) give

$$\frac{ww^*}{(\tau^2 + w\rho_2)(\tau^{*2} + w^*\rho_2^*)} = r^2$$

or
$$ww^*(1 - r^2\rho_2\rho_2^*) - r^2\rho_2^*\tau^2w^* - r^2\rho_2\tau^{*2}w - r^2|\tau|^4 = 0 \quad (21)$$

This is the general equation of a circle of radius given by

$$(r')^2 = \frac{r^4\rho_2\rho_2^*|\tau|^4}{(1 - r^2\rho_2\rho_2^*)} + \frac{r^2|\tau|^4}{(1 - r^2\rho_2\rho_2^*)}$$

Noticing that $\rho_2\rho_2^* = |\rho_2|^2$ and using eqn. (14), we have that

$$r' = r\left(\frac{1 - |\rho_2|^2}{1 - r^2|\rho_2|^2}\right) \quad . \quad . \quad . \quad . \quad (22)$$

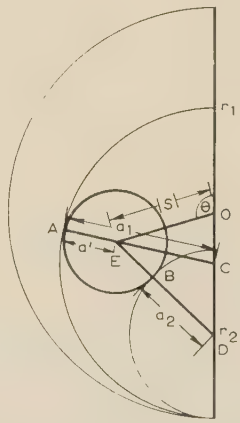


Fig. 11.—Impedance chart used for comparison of usual formula with exact formula for attenuation.

The centre of the circle in the ρ-plane is at S (Fig. 11), and is given by

$$s = \rho_1 + \frac{r^2\rho_2^*\tau^2}{1 - r^2|\rho_2|^2}$$

Hence, from eqns. (14) and (15) we have

$$s = |\rho_2|\varepsilon^{+j\phi_1}\left(\frac{1 - r^2}{1 - r^2|\rho_2|^2}\right) \quad . \quad . \quad . \quad (23)$$

Eliminating $|\rho_2|$ from eqns. (22) and (23) we have

$$r^2 - \left[\frac{1 + (r')^2 - |s|^2}{r'}\right]r + 1 = 0 \quad . \quad . \quad (24)$$

The solution of this equation is

$$r = x \pm \sqrt{(x^2 - 1)} \quad . \quad . \quad . \quad . \quad . \quad (25)$$

where
$$x = \frac{1 + (r')^2 - |s|^2}{2r'} \quad . \quad . \quad . \quad . \quad (26)$$

The choice of signs in eqn. (25) can be settled as follows:

If
$$|s| = 0,\ x = \frac{1 + (r')^2}{2r'} \geqslant 1$$

But $r \leqslant 1$, and so the negative sign must be used:

$$r = x - \sqrt{(x^2 - 1)} \quad . \quad . \quad . \quad . \quad (27)$$

Now let $x = \cosh y$; eqn. (27) becomes

$$r = \varepsilon^{-2\alpha l} = \cosh y - \sinh y = \varepsilon^{-y}$$

Hence $\qquad 2\alpha l = y = \text{arc cosh } x$

or $$2\alpha l = \text{arc cosh}\left[\frac{1 + (r')^2 - |s|^2}{2r'}\right] \quad . \quad . \quad . \quad (28)$$

It can also be shown that, for small displacements of the centre of the ρ-circle, the solution of eqn. (24) can be written as

$$r \simeq r'\left[1 - \frac{|s|^2}{1 - (r')^2}\right] \quad . \quad . \quad . \quad . \quad (29)$$

(9.2) Comparison of Usual Formula with Exact Formula for Attenuation

Consider a simplified form of Smith chart as shown in Fig. 11. From the triangle OED

$$(a' + a_2)^2 = |s|^2 + \text{OD}^2 + 2|s|\text{OD} \cos\theta \quad . \quad (30)$$

Now point D is given by[24]

$$\left(-\frac{r_2}{r_2 + 1}, 0\right) \quad . \quad . \quad . \quad . \quad . \quad (31)$$

and $$a_2 = 1/(1 + r_2) \quad . \quad . \quad . \quad . \quad (32)$$

Hence $$\text{OD} = r_2/(1 + r_2) \quad . \quad . \quad . \quad . \quad . \quad (33)$$

Substituting eqns. (32) and (33) in eqn. (30) and solving the resulting equation gives

$$r_2 = \frac{1 + 2a' + (a')^2 - |s|^2}{1 + 2|s| \cos\theta + |s|^2 - (a')^2} \text{ or } -1 \quad (34)$$

Similarly by considering the triangle OEC

$$r_1 = \frac{1 + (a')^2 - |s|^2 - 2a'}{1 + 2|s| \cos\theta + |s|^2 - (a')^2} \quad . \quad . \quad . \quad (35)$$

Hence, from eqns. (34) and (35), we have

$$\frac{r_1}{r_2} = \frac{1 - 2a' + (a')^2 - |s|^2}{1 + 2a' + (a')^2 - |s|^2} \quad . \quad . \quad . \quad (36)$$

The normal impedance-circle formula [eqn. (7)] for attenuation gives

$$\cosh 2\alpha l = \frac{1 + r_1/r_2}{1 - r_1/r_2} \quad . \quad . \quad . \quad . \quad (37)$$

Substituting the expression for r_1/r_2 in eqn. (36) gives

$$2\alpha l = \text{arc cosh}\left[\frac{1 + (a')^2 - |s|^2}{2a'}\right] \quad . \quad . \quad (38)$$

This equation is independent of the angle θ and is identical with the exact formula for attenuation [eqn. (28)] which was obtained previously.

It is evident from eqns. (34) and (35), however, that the usual formula for impedance,

$$Z_0 = Z_0'(r_1 r_2)^{1/2} \quad . \quad . \quad . \quad . \quad (39)$$

is dependent on the angle θ, indicating that the value obtained for Z_0 is influenced by the magnitude of the discontinuity through which it is being measured.

(9.3) Representation of Junction Discontinuity by a T-Network

Consider Fig. 1(*b*). The input impedance when the circuit is terminated by an impedance z_l is

$$z_i = z_{11} - \frac{z_{12}}{z_{22} + z_l} \quad . \quad . \quad . \quad . \quad (40)$$

The reflection coefficient at terminals 1–2 is then

$$\rho = \frac{z_i - 1}{z_i + 1} \quad . \quad . \quad . \quad . \quad . \quad (41)$$

Eqns. (40) and (41) give

$$\rho = \frac{z_{11} - 1}{z_{11} + 1} - \frac{2z_{12}^2/(z_{11} + 1)^2}{[z_{22} + z_l - z_{12}^2/(z_{11} + 1)]} \quad . \quad (42)$$

If we let $$z_{22} - z_{12}^2/(z_{11} + 1) = r' + jx' \quad . \quad . \quad . \quad (43)$$

and $$z_l = r_l + jx_l \quad . \quad . \quad . \quad . \quad . \quad (44)$$

eqn. (42) becomes

$$\rho = \frac{z_{11} - 1}{z_{11} + 1} - \left[\frac{2z_{12}^2}{r'(z_{11} + 1)^2}\right] \Big/ \left[\frac{r_l}{r'} + 1 + \frac{j(x_l + x')}{r'}\right] \quad (45)$$

Now, suppose z_l is modified by adding a reactance jx'. Let the modified values of the new z_l be r'' and x'', where

$$r_l/r' = r'' \text{ and } (x' + x_l)/r' = x'' \quad . \quad . \quad (46)$$

If we define a new reflection coefficient corresponding to $r'' + jx''$

i.e. $$\rho'' = \frac{r'' + jx'' - 1}{r'' + jx'' + 1} \quad . \quad . \quad . \quad . \quad (47)$$

eqns. (45), (46) and (47) provide the relationship

$$\rho = \frac{z_{11} - 1}{z_{11} + 1} - \frac{z_{12}(1 - \rho'')}{r'(z_{11} + 1)^2} \quad . \quad . \quad . \quad (48)$$

Therefore, since the z's and r' are constants, there is a linear relationship between ρ and ρ''.

Now $\qquad z_l = r_l + jx_l$, which becomes

$$z_l = r'r'' + jr'x'' - jx' \quad . \quad . \quad . \quad . \quad (49)$$

when values from eqn. (46) are substituted. If x' and r' are known, the true value of the load impedance, as measured on the ρ''-plane can be obtained.

Evaluation of x' and r'.

If $r_l = 0$ and $x_l = 0$, eqn. (46) gives
$r'' = 0$, and $x'/r' = x'' = a$, say
if $r_l = 0$ and $x_l = 1$,
$r'' = 0$, and $x'' = (x' + 1)/r' = b$, say.

Hence $$x' = a/(b - a) \text{ and } r' = 1/(b - a) \quad . \quad . \quad (50)$$

Thus the problem is resolved to the location of the points when $r_l = 0$ and $x_l = 0$ (B, Fig. 12), and when $r_l = 0$ and

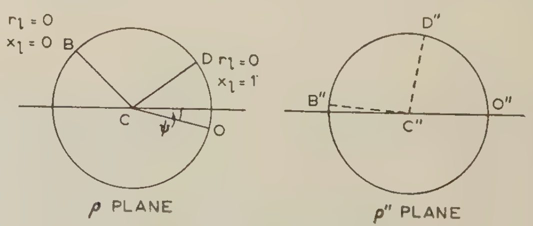


Fig. 12.—Diagram illustrating evaluation of x' and r'.

$x_l = 1$ (D on the ρ-plane). These are then transformed to the points B″ and D″ on the ρ''-plane in the usual way, giving values for a and b. Hence, values for x' and r' can be calculated from eqn. (50).

Point B is that which corresponds to a short-circuited junction. Point D is determined as follows: For a short-circuited line $Z_1 = \tanh j\beta l$

Hence $$Z_1 = j1 \text{ and } l = \lambda/8 \quad . \quad . \quad . \quad . \quad (51)$$

D is determined, therefore, when the impedance of a short-circuited line, $\lambda/8$ long, is measured through the junction.

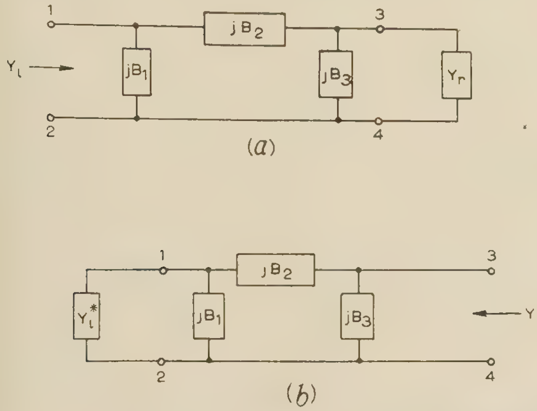


Fig. 13.—Four-terminal network for proof of conjugate-admittance theorem.

(9.4) Conjugate-Admittance Proof

Consider the four-terminal network shown in Fig. 13(a).

The three parameters B_1, B_2 and B_3 can be chosen to specify completely any lossless, four-terminal network. The input admittance when the network is terminated by an admittance Y_r is given by

$$Y_i = jB_1 + \frac{1}{1/jB_2 + 1/(jB_3 + Y_r)} \quad . \quad . \quad . \quad (52)$$

Hence $$Y_r = \frac{-(B_1B_2 + B_2B_3 + B_1B_3) + jY_i(B_2 + B_3)}{Y_i - j(B_1 + B_2)} \quad (53)$$

Now consider the case when terminals 1 and 2 are closed by an impedance Y_i^* and terminals 3 and 4 are left open [Fig. 13(b)]. The admittance presented at terminals 3 and 4 is then

$$Y_2 = jB_3 + \frac{1}{1/jB_2 + 1/(jB_1 + Y_i^*)}$$

or $$Y_2 = \frac{-(B_1B_2 + B_2B_3 + B_1B_3) - jY_i^*(B_2 + B_3)}{Y_i^* + j(B_1 + B_2)} \quad (54)$$

From eqns. (53) and (54) we see that

$$Y_2 = Y_r^* \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (55)$$

Thus, for any four-terminal network, if the input admittance presented at terminals 1 and 2 is Y_i when the network is terminated by an admittance Y_r at terminals 3 and 4, the input admittance at terminals 3 and 4 when terminals 1 and 2 are closed by an admittance Y_i^* is Y_r^*.

This information can be used to transform measured values of input admittance Y_i to corresponding values of load admittance using the following Smith-chart technique (Fig. 14).

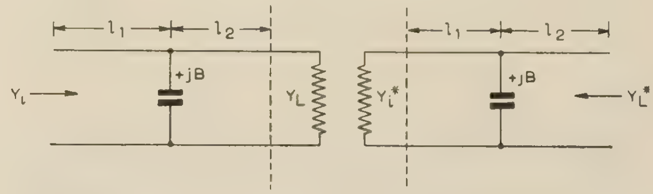


Fig. 14.—Transference of measured input admittance to corresponding cable admittance.

(a) Locate Y_i. (b) Hence find Y_i^*. (c) Move through l_1/λ away from the generator. (d) Determine Y at this point. (e) Add shunt susceptance jB. (f) Locate the new point. (g) Move through l_2/λ away from the generator. (h) The new point is Y_L^*. (i) Hence find Y_L normalized with respect to line l_2.

681.142

The Institution of Electrical Engineers
Paper No. 2693 M
Sept. 1958

SHORT-CUT MULTIPLICATION AND DIVISION IN AUTOMATIC BINARY DIGITAL COMPUTERS

With Special Reference to a New Multiplication Process.

By M. LEHMAN, B.Sc., Ph.D., Associate Member

(*The paper was first received 9th March, in revised form 26th August and 6th November*, 1957, *and in final form 31st March*, 1958.)

SUMMARY

The paper considers the application of analogues of the well-known decimal short-cut multiplication and division methods, to the control of such operations in automatic binary digital computers.

After demonstrating that the simple binary short-cut process leads, on the average, to a slowing down of multiplication, the paper developes a new process, termed the modified-short-cut (m.s.c.) process. This is defined in terms of symbolic equations and is shown to reduce the average number of additions or subtractions required during the execution of a multiplication by more than 17% to some $m/3$ (for an m-bit number), the maximum number by nearly 50% to $(m+2)/2$ and the average number of shifts by 30%.

Following a discussion of the properties of the process, and, in particular, its application to signed multiplication, comparisons are made between the relative average multiplication times of a number of alternative systems including those using circuits based on 'carry' storage or on carry-propagation-detection circuits, or on both.

Brief consideration is then given to some aspects of division in automatic binary machines. The discussion is confined to a consideration of what are believed to be ideas not previously published. It is shown, in particular, that short-cut procedures can be easily and cheaply incorporated in machines using either restoring or non-restoring division techniques.

---

(1) INTRODUCTION

The inclusion of direct multiplication and division orders in the operation codes of general-purpose, automatic, binary digital computers is now quite general; that of the former is, in fact, universal. Execution times for such orders in fixed-point machines generally range between, say, 5 and 40 times that of 'faster' orders such as addition, the exact ratio depending not only on the techniques used but also on such factors as the order-code structure (number of addresses associated with each order), the word length (of multiplier or dividend) and others. As a direct consequence of these comparative speeds and possibly also as a heritage from the allied field of desk-machine computing, typical programmes for automatic machines tend to be based on computational techniques which minimize the use of multiplication and division relative to, say, additions and shifts. Analysis of such programmes then usually indicates that any improvements in multiplication and division techniques cannot be expected to lead to significant reductions in the machine time required to solve typical problems. This analysis is by itself clearly false since speeding up the slower orders permits the use of mathematical techniques which themselves may further reduce programme lengths considerably. Thus, for example, numerical methods based on matrix algebra or continued fractions, completely alter the relative frequencies of the various orders. Furthermore, users of machines in 'real-time' problems are interested in any decrease in programme time and not only in the proportional decrease. Hence it is appropriate to consider new methods of decreasing operation times and, more particularly, various methods for controlling multiplication and division processes in binary digital computers.

Speed-up techniques in the form of short-cut processes are familiar to all users of decimal desk calculators, and their extension to the control of binary computers has recently been considered.[1] The present paper develops a new* multiplication method tentatively named the modified short-cut (m.s.c.) process, expressing its rules of operation in symbolic form. This permits a rigorous examination of its various properties and hence quantitative speed and cost comparisons between various alternative multiplication processes. In particular, it is shown that, for parallel machines, the m.s.c. process is in general to be preferred to a process based on carry storage, due to Burks, Goldstine and Von Neuman and others.[2]

The extension of short-cut techniques to various division processes is then also briefly discussed.

(2) MULTIPLICATION

(2.1) General Considerations

In conventional mechanized multiplication methods the action taken in any one cycle of a binary multiplication process is based on an examination of the 'current' multiplier bit b_t. It has been shown[3] that, in a parallel arithmetic unit (a.u.), the partial-product shift associated with each multiplication cycle may be, where an addition is also required, an integral part of the transfer normally part of the latter process or may be completely separated. In either case, the normal situation will be that the time occupied by the addition process (including carry-propagation) will be at least equal to, or substantially more than, that required for the shift. It is therefore desirable to minimize the number of addition operations required, thereby reducing the total average multiplication time. A reduction in the number of separate shifts may also prove advantageous if it can lead to a reduction in the number of cycles required for each multiplication.

In binary multiplication a short-cut method analogous to the well-known decimal method has been used to simplify the organization of signed multiplication,[4] but this does not, on the average, achieve any reduction in the total number of additions required during the course of a multiplication. The method requires that an addition or subtraction of the multiplicand from the partial product be performed whenever b_t and its right-hand neighbour b_{t-1} are unequal, the whole procedure being a mechanization of the identity:

$$\sum_{t=n}^{m} 2^t = 2^{m+1} - 2^n \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.

The paper forms part of a thesis submitted for the degree of Doctor of Philosophy at London University.

Dr. Lehman was formerly with Ferranti, Ltd., and is now with the Israeli Ministry of Defence, Israel.

* Since the paper was written, the author has become aware of the independent development of the same process by Reitwiesner. It has also recently been discussed by Professor J. E. Robertson of the Digital Computer Laboratory, University of Illinois, who quotes an earlier (1951) verbal description by D. J. Wheeler, now of the Cambridge Mathematical Laboratory.

On the usual assumption that all numbers are equally likely* (i.e. random) and that the right-hand neighbours b_t $(t < 0)$ of the least significant multiplier bit b_0 are uniquely defined (see next Section):

$$Pr\,(b_t \neq b_{t-1}) = \tfrac{1}{2} \quad t > 0 \quad . \quad . \quad . \quad . \quad (2)$$

From identity (1) it follows that a binary short-cut multiplication process may build up the product during $m + 1$ cycles (for an m-bit multiplier). Additions and subtractions (henceforth termed 'operations') alternate and are possible in each and every cycle, the maximum number of such operations being $m + 1$. From expression (2) it follows that the average number of operations will be $(m + 1)/2$ and these figures may be contrasted with the m maximum and $m/2$ average number of additions required by more standard multipliers. These inefficiencies of the short-cut process are of no real consequence in such serial machines that require all multiplications to have the same duration. Where speed is important, such machines can, for example, base multiplication on simultaneous addition of integral multiples of the multiplicand in a number of adders whose inputs are controlled by several neighbouring multiplier bits. In other types of serial machines, however, and in parallel machines it may be desirable to minimize the number of 'operations' that can occur, but the short-cut process is seen to be a retrograde step.

The source of the weakness of binary short-cut multiplication is not difficult to locate. It stems in the first place from occurrences of 'isolated' bits in the forms '. . . 0100. . ." and '. . . 1011. . .', in positive and negative multipliers, respectively, and more generally from the occurrences of oscillating sequences of form '. . . 1010. . .'. If, for example, the first combination occurs in a positive multiplier, the method leads to two operation in successive cycles, a subtraction followed by a shift and an addition where only a single addition is required in a straightforward multiplication process. Moreover, the short-cut process does not take full advantage of the binary representation. For consider an isolated zero '. . .1011. . .' in a positive multiplier. This constitutes a separation between two sequences of type (1). When short-cutting, an addition (end of one sequence) would be followed by a subtraction (beginning of next sequence). It is, however, sufficient to perform a single subtraction since

$$\sum_{\substack{t=n \\ \neq i}}^{m} 2^t = 2^{m+1} - 2^i - 2^n \quad m > i > n \quad . \quad . \quad (3a)$$

i.e. since

$$-2^{i+1} + 2^i = -2^i \quad . \quad . \quad . \quad . \quad (3b)$$

In general, short-cutting leads to redundant operation for *all* occurrences of isolated bits (and 'oscillating' sequences) in both positive and negative multipliers.

It will now be clear that it is possible to follow a modified multiplication procedure which so interprets the multiplier bits that the total number of operations that occur during a multiplication are minimized, short-cutting being suppressed whenever it is advantageous to do so. The rules for such a process (m.s.c.) are developed in the next Section, and subsequent Sections will consider its properties.

(2.2) The Modified Short-Cut Process (M.S.C.)

(2.2.1) Formal Definitions.

In order to permit an assessment of any multiplication process and to simplify its mechanization, it is useful to define its laws of procedure by means of symbolic equations. These may, for example, define a recoding of the multiplier in which '1's occur in those bit positions, relative to the original form, in which, during a mechanized process, operations are required, other bits being recorded as zero. If the particular process may require both additions and subtractions (as for example in short-cut multiplication), these may then be specified by the allocation of an appropriate sign to the non-zero bits of the coded form, the allocation of a negative sign designating, say, a subtraction. Such a recoding expresses the binary multiplier in a ternary form, but this may itself be expressed through the use of binary symbols. That is

$$\sum_{t=0}^{m-1} b_t 2^t = \sum_{t=0}^{m} T_t 2^t = \sum_{t=0}^{m} (-1)^{S_t} C_t 2^t$$

$$T_t = 0, \pm 1$$

$$S_t = 0, 1$$

$$c_t = 0, 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

For simplicity it has been assumed that the multiplier is a positive integer. The extensions of this work to numbers lying in other positive ranges, is trivial. That to negative numbers is considered for the case of the m.s.c. system, in Section 2.2.3.

The ternary recoding expression (4) is quite general. It expresses, for example, the ordinary short-cut process outlined in the previous Section if C_t and S_t are defined by expressions such as

$$C_t \equiv (b_t \neq b_{t-1}) \quad . \quad . \quad . \quad . \quad . \quad (5a)$$

$$S_t \equiv b_t \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5b)$$

This recursive definition is complete if it is recognized that in all machines $b_t\,(t < 0)$ is always implicity defined in a manner that depends on the negative-number representation adopted. Thus in a machine using the 'complements to two' or 'modulus and sign' systems $b_t\,(t < 0)$ is always zero, whereas in a '1's complements machine such bits are equal to the sign bit.

For isolated bits b_t of form $[(b_t \neq b_{t-1})\ \&\ (b_t \neq b_{t+1})]$ or for the sequences of such bits which form an oscillating sequence, pairs of non-zero C_t and C_{t+1} occur in the 'short-cut' recoding. These are clearly of opposite sign, and from identity (3*b*), the second (more significant) is redundant. This redundancy may be removed by modifying the definition (5*a*) to a form

$$C_t \equiv (b_t \neq b_{t-1})\ \&\ \bar{C}_{t-1} \quad . \quad . \quad . \quad . \quad (6)$$

The consequent redefinition required for S_t is considered below. It has been shown[1,5] that no further improvements to the short-cut process are possible (see Section 2.2.5), and hence eqn. (6) is adopted as one of the defining relations of the m.s.c. process.

It follows immediately from statements (1) and (6) that the maximum number of operations that can occur during any one multiplication is $[(m + 2)/2]$. For practical values of m this effectively equals the average number of operations in a straightforward or simple short-cut multiplier.

In order to redefine S_t, to correct the error introduced by the change in definition of C_t, it is convenient to consider a sequence concept. Continuous sequences of '0's or '1's in binary number are termed '0' and '1's sequences, denoted by $S = 0$ and $S = 1$, respectively, where S is associated with the whole sequence. The rules for the standard short-cutting process may then be formulated in terms of this sequence concept, since it follows from definition (5) that a subtraction is always required on entry into a '1's sequence and an addition upon exit from a '1's into a '0' sequence.

For the m.s.c. process the sign of those operations that arise from the occurrence of an isolated bit must be reversed, i.e. the sequence concept must be redefined so that an isolated bit (or sequences of such bits in an 'oscillating' sequence) does not alter S. An expression which defines S_t according to the

* The assumption of equal likelihood of all in-range numbers in a digital computer may not be altogether realistic. However, it is the only useful assumption which permits analysis of various multiplication methods and is probably justified where a machine is being used to its full accuracy.

above analysis, and which therefore denotes the sign of operations for the m.s.c. process, is

$$S_t \equiv \{(b_{t+1} \& b_t) \cup [(b_{t+1} \neq b_t) \& S_{t-1}]\} \quad . \quad (7)$$

This expresses the requirement that S changes only when at least two neighbouring multiplier bits b_{t+1} and b_t are equal and each unequal to b_{t-1} *and* the previous state S_{t-1}. It shows further that in the worst case (for oscillating sequence beginning at the least significant end of the multiplier) S_t may be a function of all the right-hand neighbours of the current-multiplier bit b_t.

Now C_t is non-zero only for isolated bits or when $S_t \neq S_{t-1}$, and in the latter case S_t is non-zero only if b_{t+1} is non-zero. An alternative but equivalent definition for S_t is thus

$$S_t \equiv [(C_t \& b_{t+1}) \cup (\bar{C}_t \& S_{t-1})] \quad . \quad . \quad . \quad (8)$$

In a mechanized process it is clearly of no consequence how S_t is defined when $C_t = 0$. Hence finally S_t may be very simply defined as

$$S_t \equiv b_{t+1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

The recursive definitions (7) and (8) are once again complete by virtue of the implicit definition of b_t $(t < 0)$. Expression (9), which is incompatible with the modified sequence concept previously defined, is nevertheless, in general, the most useful definition when considering the mechanization of the process, S_t being determined by one and only one bit of the multiplier during each cycle.

It has already been observed that a fundamental property of the m.s.c. process is that [from definition (6)] operations cannot occur in successive multiplier-scan cycles t and $t + 1$. Hence an operation may be directly followed by two 1-place, or for maximum speeds, a single 2-place shift, without further examination of the multiplier bits.

(2.2.2) Mechanization.

The mechanization of the m.s.c. process as defined will require the comparison, during successive cycles, of neighbouring multiplier bits b_t and b_{t-1}. In those cycles in which they are found to be unequal, an operation is followed by a 2-place or two 1-place backward shifts of the multiplier and the partial sum. If the bits are equal, a single-place shift is performed. This scanning of the multiplier bits, includes a pair b_m and b_{m-1}. For positive multipliers, b_m is zero, and it is shown in the next Section that b_m should be '1' for all negative multipliers. Hence b_m may be identified with the conventional sign bit. The need for this last scan follows immediately from identity (1), the latter implying that no further shifts occur in this cycle.

In order that the sign of each operation may be determined, a single-bit store conveniently termed S, must be provided, unless definition (9) is used as the defining equation when the necessary information can be extracted from b_{t+1}. Even in this case, it may, however, still be of advantage to provide the store S, permitting a speeding-up of the process. The setting and resetting of S may be based on any of the equations that have been determined for S_t or on a number of further alternatives. Switching between addition and subtraction in a parallel a.u. is generally obtained by complementing one of the operands and the change of state of S may be used to initiate a complementing of the multiplicand register for this purpose.

S may be set up at the commencement of each cycle t in which an operation is to be performed, by being set equal to b_{t+1} as from definition (9). Where an operation is automatically followed by two single place shifts, it is sufficient to set S equal to b_{t+2} during each cycle in which *no* operation is performed. Alternatively, or if two place-shift links are provided, S may be set equal to b_{t+3} towards the end of any cycle in which an operation has been performed and set equal to b_{t+2} in those cycles in which no operation is required. This last alternative leads to the fastest multipliers, since the addition circuits are allowed the maximum settling time before the next operation may be required.

The precise constructional requirements of an m.s.c. multiplier are naturally a function of the circuit elements from which the machine is built, but in any case the additional requirements as compared with a conventional multiplier are not large. Two 1-bit stores may be required to record S_t and C_t in each cycle. A small increase is also required in decoders or decision elements. In relation to the constructional requirements of a complete machine the resultant increase is insignificant, particularly in a fully parallel machine where the m.s.c. process finds its most obvious application and where the control circuits form only a small part of the entire machine. In such a machine, however, the provision of 2-stage shift-links may prove expensive in an absolute sense, increasing the physical size of each bit stage of the a.u. by up to 10%. The resultant overall increase in machine cost is much less than this, and the provision of these links (which may also permit the speeding up of other operations) is economically quite feasible. This is rather an important point since it will be shown in Section 2.3 that the m.s.c. process may often lead to a significant speed up of multiplication only if such links are provided.

(2.2.3) Signed Multiplication.

The m.s.c. process was developed from a consideration of the application of the well-known short-cut techniques to positive binary multipliers. It will now be shown that the process, as defined, may also be directly applied to multiplication by negative multipliers represented in either of the two common representation, '1's complements or 'complements to 2'. Hence it will follow that the process facilitates signed multiplication without any necessity for corrective steps.

Consider first the product xy where x is a negative multiplier represented in 'complements to 2'. In the arithmetic unit x appears as a positive number $2^p + x$ which can also be considered as $2^{p-1} + (2^{p-1} + x)$. If, for convenience, a fractional range of numbers is considered, the m-bit argument $x_{m-1} \ldots x_0$, of x, will appear as a positive, fractional number $(1 + x)$ for which the process is known to produce the correct result. It has been assumed that multiplication commences at the least significant end of the multiplier and in the first $(m - 1)$ cycles of the multiplication, the multiplier-unit will function correctly. The final product which results from multiplication by x can then be represented by $(1 + x)y + zy$ where zy is a factor introduced by the change of the sign-bit x_m from '0' to '1', this change effecting the mth and $(m + 1)$th cycles. It has been shown[1] that when the multiplier is represented in 'complements to 2', $z = (-1)$. Hence $(1 + x)y + zy = (2 + xy)$ or xy according as the multiplicand is positive or negative. In the former case the sign bit of the product arises automatically during the final subtraction, since negative numbers are always assumed to be preceded by an infinite sequence of sign bits. Thus the m.s.c. process may be used without modification when the multiplier is a negative number represented in 'complements to 2'.

Now consider multiplication by x, a negative (e.g. fractional) 'm-bit plus sign' multiplier represented in '1's complements. This appears in the machine as a positive number $2 + x - 2^{-m}$ and may equally well be interpreted as a number $2 + x'$, i.e. the representation of $x' = (x - 2^{-m})$ in 'complements to 2'. It has just been shown that the multiplier will produce the correct product when multiplying y by x'. But since x' is too small (i.e. too negative) by an amount 2^{-m}; the

resultant product $x'y$ will be too small by an amount $y \,.\, 2^{-m}$. Thus to obtain the true product xy it is necessary to redefine the process so as to produce an extra addition in the first (least significant) cycle when multiplying by the '1's complement of a positive number. It has been shown[1] that the process as defined by expressions (6) and (7), (8) or (9) need not be redefined if it is to include such multiplications, the desired effect arising automatically as a result of the change in the (implicit) value of $b_t (t < 0)$ from '0' to '1' in such numbers.

The validity of a direct application of the m.s.c. process to '1's complements multipliers also follows immediately from the defining equations. From expression (6) it is seen that C_t is a function only of the relationships between neighbouring bits and not of the value of individual bits. Hence the occurrences of C_t in the m.s.c. recoding of a negative number represented in '1's complements will be precisely the same as those occurring in a representation of its complement. This follows from the well-known fact that a negative number represented in this way is a copy of its modulus with all '0's and '1's interchanged for '1's and '0's, respectively. From this fact and from definition (9), say, it also follows that the values of S_t for a number and those for its '1's complement are themselves complementary. Hence the replacement of a positive multiplier by its '1's complement simply results in a change of sign for all the operations of the subsequent m.s.c. process, and this is precisely what is required to produce a product of opposite sign.

Since the sign of the multiplicand will not affect the actual multiplication process but only the 'direction' of summation, the m.s.c. process thus permits direct, signed multiplication for all combinations of multiplier and multiplicand signs.*

(2.2.4) The Average Number of Operations.

The average number of operations in the m.s.c. process may be determined from the usual assumption that all numbers are equally likely. It is shown in the Appendix (Section 7.1) that, with this assumption, the average number of operations for an m-bit plus sign multiplier is

$$\tfrac{1}{9}\left\{3m + 4 - [9 - (-1)^m](\tfrac{1}{2})^{m+1}\right\} \quad . \quad . \quad (10)$$

This number is asymptotically equal to $m/3$, so that the m.s.c. process reduces the average number of operation required by a conventional multiplier by about 17%. If two place-shift links are provided, the consequent reduction in actual multiplication time exclusive of store-access may reach 33% (compare rows a and c of Table 2A).

(2.2.5) The Optimum Process.

The development of the m.s.c. multiplication procedure has been purely intuitive, following a recognition of those effects in the standard short-cut technique which make the latter inefficient. The resultant process has been expressed by a set of expressions (6) and (7), (8) or (9) which define a ternary coded form (4) of the signed binary multiplier. It is, of course, possible to evolve a large number of alternative sets of defining equations for this general form. The question now arises as to which form, and hence which multiplication process of this type, is optimum. For parallel arithmetic units, in particular, a requirement is the minimizing of the number of operations during multiplication. Thus it is required to consider the general relationship (4) and to develop a set of defining equations for C_t and S_t in terms of b_t, subject to the restraint that ΣC_t is to be minimized. Tocher has investigated this problem[1,5] and has shown that the set of equations that arise under this restraint do, in fact, define the m.s.c. process.

* It is naturally assumed that the accumulator matches the negative number representation used. If the '1's complements representation of negative numbers is to be used because of the necessity for frequent switching between addition and subtraction, a minor difficulty results when using the m.s.c. process if the machine is only equipped with a single-length accumulator and is to produce unrounded single-length products or double-length products. The same difficulty occurs when such a machine is provided with an accumulative multiplication order, and this is due to the fact that the partial sum may change sign (several times) during the multiplication process. The alternatives of using the 'complements to 2' representation, providing a double-length accumulator or providing control equipment to introduce the necessary corrections will not be discussed here.

It should also be noted that, contrary to assertions made in Reference 1 (and also found elsewhere), the accumulator of a machine using the '1's complements representation may be initially set to either form of zero whatever the final sign of the product.

Hence this multiplication procedure leads, for any type of a.u., to the optimum multiplication method based on ternary recoding, if a minimization of the number of addition or subtraction operations is the required criterion.

(2.3) The Relative Merits of Various Multiplication Processes

Two main proposals have previously been made to speed up multiplication (and other orders) in parallel machines; the 'carry-storage' system of Burks, Goldstine, and Von Neumann[2] and the 'carry-propagation-detection' system (henceforth 'carry detection') of Gilchrist, Pomerene, and Wong.[6] A further proposal[7] combines these systems in an even faster but more complex circuit.

The m.s.c. process, in either of its forms, can clearly be used with any of these three systems to further speed up multiplication. In order to establish briefly the relative merits of the a.u.'s which arise when the various proposals are combined in different ways, assumptions must be made about the time parameters of the circuits constituting the a.u. An assessment is presented below which considers five multiplier systems as applied to six different types of machine. For convenience, the actual multiplication times (exclusive of store access) are based on figures quoted by Gilchrist, Pomerene and Wong,[6] and have, in some cases, been supplemented by a further condition that the flip-flop (or other) bit stores used have a maximum switching speed, t_2, of 0·2 microsec. Table 1 summarizes these basic times and the symbols used in Tables 2A and 2B. Of the times given in Table 1, $t_1^{5\cdot 6}$ and t_3 are quoted directly from Gilchrist *et al.*, t_1^1 is obtained from a curve reproduced in the same paper and t_1^{40} is estimated from a linear extrapolation of that curve.*

Table 1

BASIC TIME PARAMETERS

Circuit time parameters	Symbol	Time
		microsec
Carry propagation over one stage . .	t_1^1	0·12
Carry propagation over 5·6 stages (i.e. average maximum propagation time for a 40-bit number in a carry detection system)	$t_1^{5\cdot 6}$	0·21
Full-length carry propagation (for a 40-bit number)	t_1^{40}	0·90
Minimum store switching interval . .	t_2	0·20
Shift and cycle completion time	t_3	0·15

To enable assessments to be made of multiplication times of machines in different cost ranges, three variations in machine design have been considered. The first differentiates between machines in which the bit store does or does not impose a restriction on the working speed of the machine. The value

* The expressions and results quoted here differ somewhat from those given in Reference 1. Those were based on a time for t_1^1 of 0·045 microsec. This has now been modified to 0·12 microsec, since closer study of Gilchrist's results has revealed an initial non-linearity or dead-time. This affects not only the magnitudes of various time parameters, but also some of the expressions given in Table 2 since it is in places necessary to select the larger of alternative terms. The original Table 5 of Reference 1 is also misleading since the value $t_1^{5\cdot 6} = 0\cdot 21$ microsec quoted by Gilchrist has been used in all places where terms of the form $5\cdot 6t_1^1$ occur and $5\cdot 6 \times 0\cdot 045 \neq 0\cdot 21$.

Table 2A.—Typical Average Multiplication Times

		Type of machine					
		1	2	3	4	5	6
		0·2 microsec limit on reset time of flip-flops				No limit	
		Single-sided store access		Push-pull store access			
	Type of multiplier	No carry detection	Carry detection	No carry detection	Carry detection	No carry detection	Carry detection
a	Conventional	$20(t_1^{40} + 3t_2)$ 30·0	$20(t_1^{5\cdot6} + 3t_2)$ 16·2	$20(t_1^{40} + t_2 + t_3)$ 25·0	$20(t_1^{5\cdot6} + t_2 + t_3)$ 11·2	$20(t_1^{40} + 2t_3)$ 24·0	$20(t_1^{5\cdot6} + 2t_3)$ 10·2
b	M.S.C. with single place shift-link	$13\cdot8t_1^{40} + 68\cdot2t_2$ 26·0	$13\cdot8t_1^{5\cdot6} + 68\cdot2t_2$ 16·5	$13\cdot8(t_1^{40} + t_3) + 27\cdot2t_2$ 19·9	$13\cdot8(t_1^{5\cdot6} + t_3) + 27\cdot2t_2$ 10·4	$13\cdot8t_1^{40} + 41t_3$ 18·6	$13\cdot8t_1^{5\cdot6} + 41t_3$ 9·0
c	M.S.C. with single and double place shift links	$13\cdot8t_1^{40} + 42\cdot6t_2$ 20·9	$13\cdot8t_1^{5\cdot6} + 42\cdot6t_2$ 11·4	$13\cdot8(t_1^{40} + t_3) + 14\cdot4t_2$ 17·4	$13\cdot8(t_1^{5\cdot6} + t_3) + 14\cdot4t_2$ 7·8	$13\cdot8t_1^{40} + 27\cdot2t_3$ 16·5	$13\cdot8t_1^{5\cdot6} + 27\cdot2t_3$ 7·0
d	Conventional with carry storage	$t_1^{40} + 81t_2$ 17·1	$t_1^{5\cdot6} + 81t_2$ 16·4	$40t_1^1 + t_1^{40} + 41t_3$ 11·9	$40t_1^1 + t_1^{5\cdot6} + 41t_3$ 11·0	$40t_1^1 + t_1^{40} + 41t_3$ 11·9	$40t_1^1 + t_1^{5\cdot6} + 41t_3$ 11·0
e	Combined types *c* and *d*	$57\cdot4t_2 + t_1^{40}$ 12·4	$57\cdot4t_2 + t_1^{5\cdot6}$ 11·7	$13\cdot8(t_1^1 + t_3) + 14\cdot4t_2 + t_1^{40}$ 7·5	$13\cdot8(t_1^1 + t_3) + 14\cdot4t_2 + t_1^{5\cdot6}$ 6·8	$13\cdot8t_1^1 + 28\cdot2t_3 + t_1^{40}$ 6·8	$13\cdot8t_1^1 + 28\cdot2t_3 + t_1^{5\cdot6}$ 6·1

Table 2B.—Ratios of Average Multiplication Times Relative to Fastest Multiplier

		Type of machine					
		1	2	3	4	5	6
		0·2 microsec limit on reset times of flip-flop				No limit	
		Single-sided store access		Push-pull store access			
	Type of multiplier	No carry detection	Carry detection	No carry detection	Carry detection	No carry detection	Carry detection
a	Conventional	4·9	2·7	4·1	1·8	3·9	1·7
b	M.S.C. with single place shift-link	4·3	2·7	3·3	1·7	3·0	1·5
c	M.S.C. with single and double place shift link	3·4	1·9	2·9	1·3	2·7	1·1
d	Conventional with carry storage	2·8	2·7	2·0	1·8	2·0	1·8
e	Combined types *c* and *d* ..	2·0	1·9	1·2	1·1	1·1	1·0

given to t_2 is considered to be representative of the speeds that can be obtained from simple flip-flops. Consideration has also been given to the alternatives of providing single-sided or push-pull access to the a.u. registers, the former requiring that a register must always be reset before new information can be written in. In the latter, two sets of inter-register connections (highway and anti-highway) are provided; these carry signals corresponding to '1's and '0's, respectively, and in any interchange of information in the a.u., one or other signal will be present in each bit stage of the parallel machine. Hence registers need not be reset before new information is written in. Finally, comparisons are made between machines incorporating carry detection and those that do not, and in all cases '40-bit plus sign' numbers have been considered.

In attempting to assess the merits of the various combinations summarized in Tables 2A and 2B, it must be borne in mind that all the results are fairly sensitive to relative changes in the magnitudes of the various parameters. The times quoted for units not using carry-detection circuits do not include safety margins, and hence the speed gain resulting from the use of such circuits is somewhat greater than indicated. On the other hand, the final times and ratios do not include store access-times. The relative effect of these depend not only on their value, but also on the order code (being minimized in a code which minimizes reference to the store[8]) and on the logical structure of the machine. It is thus difficult to include consideration of access times in the Tables, but the question is briefly considered at the end of this Section.

The conventional multiplier (*a*) requires no further comment. Case (*b*) considers a multiplier which requires only a small amount of extra control equipment and is therefore not significantly more expensive than one of type (*a*). This system gives a real increase in speed, only where carry-detection circuits are not used, and in the case of machine 2 is even slightly slower than a conventional multiplier because of its requirement for an extra multiplication cycle. This is, however, balanced by the extra time that the conventional multiplier requires to produce a signed product.

The introduction in case (*c*), of the extra link for the partial-product and multiplier backward-shift facility, is seen to produce a speed-up in all cases, since the average number of multiplication cycles is reduced by one-third. The provision of this additional shift-facility will increase the size of a typical a.u. by some 10–15%, but this is not serious since the consequent increase in size of the whole machine may be expected to be less than 5% and the increase in cost even less. These and all other estimates given are naturally only very approximate since they clearly depend on the type of logic and circuits used in the machine as well as on its final size. In all probability the type of machine for which the speed-ups quoted are of consequence will, in any case, be a large machine, and the expected increases in cost of the whole installation will be much less than the figure quoted.

Case (*d*) considers a machine based on the carry-storage concept. The provision of such a carry store increases the size of a conventional parallel a.u. by some 25%, whereas a carry-detection circuit will, at the most, increase its size by, say, 15%. Even with a conventional multiplier the carry-detection circuit is clearly more efficient, producing a faster multiplier for a smaller investment. Moreover, carry detection will speed up all two-operand arithmetic operations, whereas carry storage is useful only in multiplication, or, if this is included in the order code, in accumulative addition. Thus carry storage is by itself not the most useful innovation. An exception to this conclusion may, however, arise in a parallel machine whose electronic circuits are to be based on circuits (packages) initially developed for use with serial machines, or for circuits using magnetic cores for logical operations. In such machines it may happen that each circuit possesses an inherent storage property, and in this case, carry storage may be obtained with little further cost.

Carry storage may be combined with either of the m.s.c. multipliers. The advantages of such multipliers arises from the minimization of additions and possibly also of shifts. Hence it is clear that in a combined multiplier the carry register should have shifting facilities to permit the restriction of additions of carries to those cycles in which an addition (of the multiplicand or its complement) is in any case required. Full advantage may then be taken of the m.s.c. system, but such a scheme also demands that during multiplication the carry output of the addition network feeds one place back. At the conclusion of multiplication, when the carries have finally to be dissipated (as during an ordinary addition), they feed one place forward. Thus the carry output of the arithmetic circuits must be switchable to at least two destinations, and the system (*e*) becomes very large and complex requiring, in all cases, at least 25% more equipment than the corresponding requirements for (*c*). In comparing (*c*) and (*e*) it is necessary to consider only cases already involving carry-detection circuits, since the result that carry detection is to be preferred to carry storage holds also for the complex system. For these cases (*e*) is seen to be slower than (*c*) for a 'simple' machine and only some 12% faster in the other two cases. Since this small advantage is even further reduced when overall multiplication times including store access and overall programme times are considered, it seems fair to conclude that (*e*) does not form the basis for a reasonable system with most present-day circuits.

Turning now to the alternative machines, it is clear that the provision of carry detection circuits provides important speed increases of the order of two, for multiplication. The circuit will enlarge the a.u. by some 10%, and this is approximately the same as the cost of replacing single-sided access to the flip-flops or other registers, by push-pull access. Since the latter does not produce the same advantages it is clear that carry detection is, in general, to be preferred. However, the two systems are to some extent complementary, since carry-detection circuits are essentially symmetrical in the operands and their complements. Hence the combined cost of installing both features is somewhat less than their individual incorporation, and it will often be advantageous to use both.

With the particular set of values selected for the various parameters, the effect of removing restrictions on the working rate of the (flip-flop) bit-store is not pronounced and appears to lead to a speed-up of only some 10%. In practice, however, the magnitude of this parameter is of vital importance since it determines whether time control is required within the a.u. Thus it is operationally desirable, and in an asynchronous machine essential, that $t_2 \ll (t_1^1 + t_3)$.

Mention has been made of a second advantage of the m.s.c. process, that of reducing maximum multiplication times. Computation of such times shows that, for the conventional process, maximum times are some 50% greater than average times, whereas for the m.s.c. process this figure is reduced to 10%. In the pure carry-storage system multiplication always takes the same time, whereas, for the combined system, the excess of maximum over average times is only some 2%. Hence, from the point of view of maximum performance times for complete programmes, there is no significant difference between the more complex systems, but each has some slight advantage over conventional multipliers in such machines that permit variable-length operations.

The effect of store access times will, of course, be to reduce the real gains to be expected from any of these multiplication methods by an amount which depends on the magnitude of

various store characteristics, on machine logic and on the order code to be used. Present-day magnetic-core stores have access (cycle) times of about 5 microsec, and 1 microsec stores appear to be under active development. Hence the fast multiplication times quoted are meaningful and do represent a measure of the effect on programmes of improved methods of multiplication.

It may be concluded that the m.s.c. process, whether combined with carry detection or not, has real advantages that should lead to its adoption for the control of multiplication in parallel and certain types of serial machines. The advantages of the carry-storage system appear to be confined to parallel machines using circuit elements which all involve a settling, sampling or delay time which is of the order of one bit-time.

(3) DIVISION

The standard techniques of restoring and non-restoring division, as applied to binary machines, are well known,[1,2,4,5,9] and the discussion in the present paper will be restricted to such aspects, particularly of short-cut processes, which are considered new or which can be presented in a new or improved form.

The main problem in non-restoring division processes is in the manner of recording the quotient bits. In every cycle of the standard process either an addition or subtraction operation is performed. To record the quotient it would thus appear necessary to attach a significance of $(+1)$ or (-1) to each of its bits. It is, of course, desirable to keep the circuit elements of the machine as simple and consistent as possible and hence to record the quotient in terms of the standard '0' and '1'. That this can be done has previously been shown,[1,2,4] and a very simple proof is given in the Appendix. From this proof it follows that the signed and rounded quotient may easily be obtained from the 'coded' version, which results from the insertion of a '1' in the least-significant bit position of the pseudo-quotient whenever a subtraction occurs, and the insertion of a '0' when the operation performed is addition. A forward shift of the resultant partial quotient is, of course, required in each cycle of the process, and the leading bit of the pseudo-quotient should at the conclusion of division appear in the sign position of the number. It is then sufficient to complement this leading bit and to set the least-significant bit unequal to the resultant sign bit.

It is possible to improve either of the two division techniques by the application of simple tests which lead to a form of short-cutting. For the restoring process, trial subtractions should be suppressed if simple tests on the leading bit or bits of the operand arguments permit the prediction of the result of such a subtraction.[1,9] In the case of the non-restoring process an operation (addition or subtraction) may be suppressed when similar tests indicate that the partial remainder would change sign if such an operation were performed.[1] The test required in either case implies, in its simplest form, that, with the usual convention, the divisor D must always be in a standard form (say $1 > D \geqslant \frac{1}{2}$), the partial remainder must also be in this range before any operation need be performed. Such a test for standard form will always be provided in floating-point machines, and generally also in fixed-point machines, where it is very useful to have a special 'standard-form shift' order. Thus the cost of incorporating this very primitive form of short-cutting, which may produce an average division speed-up of up to some 50%, is quite negligible. More refined forms of short-cut tests have also been discussed.[1] These produce progressively smaller reductions in the average duration of the division process, as the number of bits on which the test is based is increased.

A quantitative determination of the time saving as a function of the number of bits examined has not yet been made, since this would require some assumption about the expected distribution of bits in the successive partial remainders. These are clearly not random so that the usual simple assumptions cannot be made. A knowledge of the distribution of these bits can probably be obtained most easily by performing 'synthesized' divisions in an electronic computer and counting the occurrences of, say, non-zero bits in the successive partial remainders. In the absence of relevant information, one may safely assume that the simplest form of short-cutting is certainly always justified, and a form based on a comparison of the two leading bits of the partial remainder and the divisor will also be worth while, since, even in this case, additional constructional requirements are small.

One further problem must be considered in connection with short-cut techniques as applied to non-restoring division, namely that of the modifications required in the procedure of building up the pseudo-quotient bits. The modification to the process described above requires that the normal feed connection for the quotient bit be inverted in the second cycle of a short-cut sequence. This inverse connection must be retained until after conclusion of short-cutting and the first cycle of more normal operation. This rule is clearly true if only one such short-cut cycle is performed. It has been shown[1] that, if the rule is true for an n-cycle sequence, it is true for an $(n + 1)$-cycle sequence. Hence by induction the generality follows.

Short-cut non-restoring division which yields the fastest practical division process may thus be attained by the provision of only a very small amount of additional equipment over that required for the straightforward techniques. Moreover, the extra cost of incorporating a division order in relation to the cost of the whole machine is quite small, particularly in a parallel machine. In the new machine of Reference 1, for example, special division equipment increases the estimated cost of the machine by less than 3%. Thus it appears that a division order, based on short-cut techniques, should be provided for all parallel a.u.'s. Division should be similarly provided in general-purpose serial machines, and where such machines are not restricted to fixed-length orders, these may also take advantage of short-cut techniques. This may be done by shifting the partial remainder and partial quotient in each cycle by a number of places determined by the short-cut test circuits.

(4) CONCLUSIONS

The paper has considered the application of short-cut methods to multiplication and more briefly, to division. It has shown that their application leads to a significant reduction in operation times in machines in which these need not be of predetermined (fixed) duration. A multiplication method has been developed which by modifications to more standard techniques may reduce the average multiplication time by more than 30%. The method also reduces the maximum multiplication time for an overall increase in the size of the machine certainly not exceeding 5%.

Comparisons of multiplication times, exclusive of store access, have been presented for several multiplication techniques. These have been based on somewhat arbitrary, but it is hoped, representative circuit speed characteristics. Analysis of the resultant times indicates the value of the new process.

Short-cut techniques may also be applied to the mechanization of binary division for an even smaller increase in the overall machine size. The resultant reduction in the duration of the two longer operations is a function of a large number of properties of the a.u. and of the machine, and a general estimate of the efficiencies of the processes is difficult to make, particularly in the case of division. It is believed that the techniques discussed will find their application in the logical design of high-speed

computers where their true value can be assessed only when they have been incorporated in a number of actual designs.

(5) ACKNOWLEDGMENTS

The author wishes to acknowledge many constructive discussions with Dr. K. D. Tocher, formerly of the Department of Mathematics, Imperial College of Science and Technology. He also wishes to acknowledge the contributions, made during numerous discussions, of his former colleagues in the Logical Design Group of the London Computer Laboratory of Ferranti Ltd., and in particular those of Messrs. R. W. Arnott and M. Hill.

(6) REFERENCES

(1) LEHMAN, M.: 'Parallel Arithmetic Units and their Control' (Ph.D. Thesis, University of London, February, 1957).

(2) BURKS, A. W., GOLDSTINE, H. H., and VON NEUMAN, J.: 'Preliminary Discussion of the Logical Design of an Electronic Computing Instrument' (Part I, Institute of Advanced Study, Princeton, New Jersey, 1947).

(3) TOCHER, K. D., and LEHMAN, M.: 'A Fast Parallel Arithmetic Unit', *Proceedings I.E.E.*, Paper No. 2108 M, November, 1956 (**103** B, Supplement No. 3, p. 520).

(4) BOOTH, A. D., and BOOTH, K. H. V.: 'Automatic Digital Calculators' (Butterworth Scientific Press, 1953), pp. 44–47.

(5) TOCHER, K. D.: 'Techniques of Multiplication and Division for Automatic Binary Computers', *Journal of Mechanics and Applied Mathematics* (to be published).

(6) GILCHRIST, B., POMERENE, J. H., and WONG, S. Y.: 'Fast Carry Logic for Digital Computers', *Institute of Radio Engineers Transactions on Electronic Computers*, December, 1955, p. 133.

(7) ESTRENE, G., GILCHRIST, B., and POMERENE, J. H.: 'A Note on High-Speed Digital Multiplication', *ibid.*, September, 1956, p. 140.

(8) TOCHER, K. D.: 'The Classification and Design of Operation Codes for Automatic Computers', *Proceedings I.E.E.*, Paper No. 2141 M, September, 1956 (**103** B, Supplement No. 1, p. 125).

(9) RICHARDS, R. K.: 'Arithmetic Operations in Digital Computers' (Van Nostrand, 1955), pp. 168–172.

(7) APPENDICES

(7.1) The Average Number of Operations in the M.S.C. Process

From definition (6):

$$\begin{aligned} C_t &\equiv \{(b_t \neq b_{t-1}) \;\&\; \bar{C}_{t-1}\} \\ &\equiv \{(b_t \neq b_{t-1}) \;\&\; \overline{(b_{t-1} \neq b_{t-2}) \;\&\; \bar{C}_{t-2}\}} \\ &\equiv \Big\{(b_t \neq b_{t-1}) \;\&\; \Big\{\{(b_{t-1} = b_{t-2}) \cup \{(b_{t-1} \neq b_{t-2}) \\ &\qquad \& \; (b_{t-2} \neq b_{t-3}) \;\&\; \{(b_{t-3} = b_{t-4}) \ldots \\ &\qquad \ldots \cup \{(b_4 \neq b_3) \;\&\; (b_3 \neq b_2) \;\&\; \{(b_2 = b_1) \cup \{(b_2 \neq b_1) \\ &\qquad \& \; (b_1 = b_0) \;\&\; (b_0 = b_{-1})\Big\} \ldots \Big\} \quad t \text{ odd} \quad (11a) \end{aligned}$$

$$\begin{aligned} \text{or} \quad &\ldots \cup \Big\{(b_3 \neq b_2) \;\&\; (b_2 \neq b_1) \;\&\; \{(b_1 = b_0) \cup \{(b_1 \neq b_0) \\ &\qquad \& \; (b_0 \neq b_{-1})\} \ldots \Big\} \quad t \text{ even} \quad (11b) \end{aligned}$$

It should be noted that definition (6) is expanded so as to express all 'or' connectives exclusively.

From the condition that all numbers are equally likely it follows that

$$Pr(b_t \neq b_{t-1}) = Pr(b_t = b_{t-1}) = \tfrac{1}{2} \quad . \quad . \quad (12)$$

Hence

$$\begin{aligned} \underset{t \neq m}{Pr(C_t = 1)} &= \tfrac{1}{2}\Big(\tfrac{1}{2} + (\tfrac{1}{2})^2\big(\tfrac{1}{2} + \ldots (\tfrac{1}{2})^2(\tfrac{1}{2} + (\tfrac{1}{2})^3)\big) \ldots \Big) \; t \text{ odd} \\ &= \tfrac{1}{2}\Big(\tfrac{1}{2} + (\tfrac{1}{2})^2\big(\tfrac{1}{2} + \ldots (\tfrac{1}{2})^2(\tfrac{1}{2} + (\tfrac{1}{2})^2)\big) \ldots \Big) \; t \text{ even} \\ &= \tfrac{1}{3}[1 - (\tfrac{1}{2})^{t+1}] \qquad t \text{ odd} \\ &= \tfrac{1}{3}[1 + (\tfrac{1}{2})^{t+1}] \qquad t \text{ even} \\ &= \tfrac{1}{3}[1 + (-1)^t(\tfrac{1}{2})^{t+1}] \\ &\qquad\qquad \text{all } t \neq m \quad . \quad . \quad . \quad (13) \end{aligned}$$

If $t = m$, then for any m, the final term in an expanded form of both eqns. (11*a*) and (11*b*) is missing, and hence

$$\underset{t=m}{Pr\,(C_t = 1)} = \Big[\underset{t \neq m}{Pr\,(C_t = 1)} - (\tfrac{1}{2})^{m+1}\Big] \quad . \quad . \quad (14)$$

whence

$$\underset{\text{all } t}{Pr\,(C_t = 1)} = \Big\{\tfrac{1}{3}[1 + (-1)^t(\tfrac{1}{2})^{t+1}] - \delta_t^m(\tfrac{1}{2})^{m+1}\Big\} \quad . \quad (15)$$

Therefore the average number of occurrences of $C_t = 1$ in the coded form $\sum_{t=0}^{m} (-1)^{S_t} C_t 2^t$, i.e. the average number of additions or subtractions in a mechanized m.s.c. multiplication is

$$\begin{aligned} &\sum_{t=0}^{m} \tfrac{1}{3}[1 + (-1)^t(\tfrac{1}{2})^{t+1}] - (\tfrac{1}{2})^{m+1} \\ &= \tfrac{1}{3}\Big[m + 1 - \sum_{t=1}^{m} (-1)^t(\tfrac{1}{2})^t\Big] - (\tfrac{1}{2})^{m+1} \\ &= \tfrac{1}{9}[3m + 4 - (-1)^{m+1}(\tfrac{1}{2})^{m+1} - 9(\tfrac{1}{2})^{m+1}] \\ &= \tfrac{1}{9}\Big\{3m + 4 - [9 - (-1)^m](\tfrac{1}{2})^{m+1}\Big\} \quad . \quad . \quad (16) \end{aligned}$$

the result previously quoted as expression (10) in Section 2.2.4.

(7.2) The Quotient in Non-Restoring (Non-Short-cut) Division

The quotient obtained from a non-restoring division process may be recorded by an m-bit sequence of '1's and '0's, the former being inserted (as in restoring processes) when a subtraction is performed, and the latter in those division cycles in which an addition occurs. Thus the quotient is expressed in a representation:

$$\begin{aligned} Q &= \sum_{t=0}^{m-1} (-1)^{(1-b_t)} 2^t \quad . \quad . \quad . \quad . \quad . \quad . \quad (17) \\ &= \sum_{t=0}^{m-1} b_t 2^t - \sum_{t=0}^{m-1} (1 - b_t) 2^t \quad . \quad . \quad . \quad (18) \\ &= x + y, \text{ say} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19) \end{aligned}$$

Q has, without loss of generality, been expressed as an integer, and it is immediately clear that it is a signed odd number lying in the range

$$|Q| \leqslant 2^m - 1$$

The computer can interpret the sequence of $m - 1$ bits which make up the quotient either as a positive number x, or as a negative number y represented in '1's complements but without a sign-bit, where x and y are defined by eqns. (18) and (19).

Now $\qquad x + y = Q \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$

also clearly $\qquad x - y = 2^n - 1 \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$

hence $\qquad Q = 2x - 2^n + 1 \quad . \quad . \quad . \quad . \quad (21)$

$\qquad\qquad = 2y + 2^n - 1 \quad . \quad . \quad . \quad . \quad (22)$

Thus the machine, which recognizes only pure binary numbers

and their complements, can obtain the true quotient Q from the 'pseudo-quotient' x or y, by a process which is formerly equivalent to eqns. (21) and (22), i.e. a shift forward, complementing the leading (now sign) bit and forcing the least-significant bit unequal to the resulting sign bit. Expressions (21) and (22) give the justification of this 'conversion' process according to whether the resultant quotient is positive (based on x) or negative (expressed in '1's complements and based on y). The subtraction or addition of the term 2^n required by eqns. (21) and (22) may be obtained simply by complementing, since, as has been noted above, the leading bit will always initially be '1' or '0' according to whether the number is positive or negative. The quotient thus obtained is rounded according to the well-known 'forcing' rule used in many machines, but an unrounded quotient can also be obtained by a slight modification of the division and correction process.

DISCUSSION ON 'THE APPLICATION OF PRINTED-CIRCUIT TECHNIQUES TO THE DESIGN OF MICROWAVE COMPONENTS'*

SOUTH-EAST SCOTLAND SUB-CENTRE AT EDINBURGH, 7TH APRIL, 1958

Mr. E. Jamieson: The author gives a very detailed account of both the theoretical and practical aspects of microwave printed circuits. In Table 1 he compares the different strip-line systems with regard to certain parameters, but I believe that this could usefully be extended. The microwave engineer is interested, not only in the absolute value of strip-line, but in its comparative value with respect to other transmission-line systems. I suggest that strip-line be considered with regard to the various major classes of microwave system, and also with respect to the frequency of operation. The main microwave systems are (*a*) primary radars, (*b*) multi-channel microwave receivers, and (*c*) microwave links.

With regard to (*a*) there would seem to be little doubt that the exclusive use of strip-line is impossible, owing to the close proximity of transmitter and receiver, with a power difference of over 100 dB, and the radiation problem of strip-line. The power-handling capacity of strip-line is also probably insufficient for any reasonably high-powered transmitter.

For (*b*), where no power-handling problems are involved, there would seem to be a prima facie case for strip-line, but if close spacing and channel-to-channel isolation of perhaps 40–60 dB is required, it is likely that interference problems will arise.

System (*c*) appears to provide a worth-while application for strip-line; no high power is involved, a single broad-band channel is required, and simple mass-production techniques pay large dividends.

For microwave frequencies it would again appear that strip-line is limited. At the lowest band (1 Gc/s), the major components—aerial, transmitter valve, etc.—are large, and coaxial line is usually adequate as regards space, and also lends itself readily to probe feeds to coaxial-line cavities, etc. Between 2 and 6 Gc/s, where large-section waveguide is normally used, there would seem to be a very good case for strip-line. Once the X-band is reached, however, the design of strip-line components is much more difficult; in particular, waveguide-to-strip-line transitions, hybrids and crystal mounts require closer tolerances.

It would thus appear that the author has chosen the ideal field for the exploitation of printed microwave circuits, and that engineers in the other parts of the mircowave field can only admire, but not imitate.

* DUKES, J. M. C.: *Proceedings I.E.E.*, Paper No. 2401 R, August, 1957 (see 105 B, p. 155).

Finally, has the author considered the use of flexible backing? I believe that C.S.F. in France has produced a very fine filter consisting of a number of feet of printed circuit, rolled up and canned.

Mr. J. M. C. Dukes (*in reply*): Table 1 could be extended, as Mr. Jamieson suggests, by the addition of the following two columns (as before, the numbers indicate order of preference):

Property	*Coaxial*	*Waveguide*
Ease of fabrication ..	6	5
Bulk or weight ..	4	5
Ease of testing ..	$\leqslant 1$	$\leqslant 1$
Ease of analysis ..	$\lessdot 1$	< 1
Dissipative losses ..	3	$\leqslant 1$
Radiation losses ..	$\leqslant 1$	$\leqslant 1$
Topological limitations	As for triplate (i.e. rather severe)	Similar to balanced line
Appropriate fields of application	Broad-band components and medium-Q-factor filters	Medium bandwidth components and high-Q-factor filters

Rectangular waveguide may be regarded as a form of balanced strip transmission line, and the properties are in many ways similar. It is likewise not easily connected to coaxial cable—a disadvantage which is accentuated by its dispersive properties. It lends itself somewhat more easily to the formation of E-plane series junctions, and has the obvious advantage that radiation can be expressed completely, or virtually so. Coaxial line resembles triplate very closely; the only practical advantage being the greater ease with which spurious mode propagation may be avoided.

Mr. Jamieson's classification of microwave systems would appear to ignore at least two important fields of interest, namely low-power airborne radar equipment and broad-band microwave receivers. Strip transmission lines may be used with advantage in both these applications. As suggested by Mr. Jamieson, it may also be used on microwave links, but so far the application has been limited to p.t.m. systems or low-channel-capacity f.m. systems. With trunk-route f.m. systems carrying 600 or more speech channels, the very low reflection coefficient essential to an acceptable level of intermodulation noise has so far tended to discourage the adoption of strip transmission line techniques.


A.T.E + MARCONI
Combined Operations
The complexity of modern radio-multichannel systems involving hundreds of telephone channels has brought about a collaboration between the two leading specialist organizations in the field—Marconi's in radio, and A.T.E. in carrier transmission.
This completely unified approach to development, systems planning, supply, installation, maintenance of equipment and training of personnel covers radio-multichannel systems in the V.H.F., U.H.F. and S.H.F. frequency bands all over the world.
Marconi
ATM
REGD TRADE MARK
Full information may be obtained from either: MARCONI'S WIRELESS TELEGRAPH COMPANY LIMITED, CHELMSFORD, ESSEX, ENGLAND.
or AUTOMATIC TELEPHONE & ELECTRIC CO. LTD., STROWGER HOUSE, ARUNDEL STREET, LONDON, W.C.2

PROCEEDINGS OF THE INSTITUTION OF ELECTRICAL ENGINEERS

Part B. RADIO AND ELECTRONIC ENGINEERING (INCLUDING COMMUNICATION ENGINEERING), SEPTEMBER 195

CONTENTS

Bibliographical References.—It is requested that bibliographical reference to an Institution paper should always include the serial number of the paper and the month and yea of publication, which will be found at the top right-hand corner of the first page of the paper. This information should precede the reference to the Volume and Part.

Example.—SMITH, J.: 'Reflections from the Ionosphere', *Proceedings I.E.E.*, Paper No. 3001 R, December, 1954 (102 B, p. 1234).

Published by The Institution, Savoy Place, London, W.C.2. *Telephone:* Temple Bar 7676. *Telegrams:* 'Voltampere, Phone, London.'
Printed by Unwin Brothers Limited, Woking and London.